

MASTERARBEIT

---

# Statistische Modellierung von Klimasimulationsergebnissen zur Berechnung der zugehörigen Strahlungswirkung

---



Sabrina Koopmans

Betreuer: Prof. Dr. Volker Schmid, LMU  
Projektpartner: Dr. Michael Ponater, DLR  
Prof. Dr. Robert Sausen, DLR

vorgelegt am 04.03.2016



## **Abstract**

For many studies in climate science a major part of the experimental concept is the comparison of two or more simulations to answer the question: Is there a difference? Depending on the simulation setup and results different procedures have been developed and put to use over the last decades. Now more than ever due to the extensive production of huge datasets, more sophisticated questions can be asked and need to be answered. This work will give an overview of the difficulties involved in providing answers, shortly reviews some of the existing methodology and proposes some methods for further study. The main part consists of the application of three selected evaluation methods, univariate t-test, principal components analysis and regression models to a specific dataset, which will illustrate the strenghts and weaknesses of each. Concluding this work, several recommendations will be given for a statistically sound evaluation strategy depending on the questions asked and datasets produced.

# Inhaltsverzeichnis

<b>1. Einführung</b>	<b>6</b>
1.1. Situationsbeschreibung . . . . .	6
1.2. Motivation . . . . .	6
1.3. Kriterien zur Methodenauswahl . . . . .	7
1.4. Verwendete Methoden . . . . .	10
<b>2. Klimamodellsimulationen</b>	<b>12</b>
2.1. Einführung in Allgemeine Zirkulationsmodelle . . . . .	12
2.2. Beschreibung des Beispieldatensatzes . . . . .	12
2.2.1. Beschreibung der Datenstruktur . . . . .	13
2.2.2. Eigenschaften der Variablen . . . . .	15
2.2.3. Wichtige Merkmale des Datensatzes . . . . .	19
<b>3. Hypothesentests</b>	<b>20</b>
3.1. Mathematischer Hintergrund . . . . .	20
3.1.1. Allgemeine Theorie statistischer Hypothesentests . . . . .	20
3.1.2. Multiples Testen . . . . .	21
3.1.3. Der Student'sche t-Test und Alternativen . . . . .	23
3.1.4. Tests auf Normalverteilung . . . . .	23
3.2. Hypothesentests mit dem Beispieldatensatz . . . . .	24
3.2.1. Test auf Normalverteilung und Auswahl der t-Testprozedur . . . . .	24
3.2.2. Vergleich der t-Tests . . . . .	25
3.2.3. Vergleich der Multiplen-Testen-Korrekturen . . . . .	25
3.2.4. Ergebnisse der korrigierten Hypothesentests . . . . .	27
3.3. Anwendungsbereich von Hypothesentests . . . . .	27
<b>4. Orthogonalzerlegung PCA (EOF)</b>	<b>29</b>
4.1. Theoretischer Hintergrund . . . . .	29
4.1.1. Beschreibung des Verfahrens . . . . .	29
4.1.2. Implementierungsmöglichkeiten des Verfahrens . . . . .	30
4.2. PCA mit dem Beispieldatensatz . . . . .	31
4.2.1. Vergleich der Verwendung von Jahresmittelwerte, gleichartigen Monaten und dem kompletten Zeitverlauf . . . . .	31
4.2.2. EOF-Analyse für die anderen Variablen . . . . .	34
4.3. ICA für den Beispieldatensatz . . . . .	35
4.3.1. ICA der Sulfat-Konzentrationsdifferenz der untersten Höhengschicht . . . . .	35
4.3.2. ICA für die anderen Variablen in der untersten Höhengschicht . . . . .	35
4.4. Vergleich PCA und ICA . . . . .	36
4.5. Anwendungsbereich PCA . . . . .	37
<b>5. Regressionsmodelle</b>	<b>38</b>
5.1. Mathematischer Hintergrund . . . . .	38
5.1.1. Lineare Modelle . . . . .	38
5.1.2. (Generalisierte) additive Modelle . . . . .	39
5.1.3. Geostatistik . . . . .	40



5.1.4.	Zeitreihenmodelle . . . . .	40
5.1.5.	Modellgütekriterien . . . . .	41
5.2.	Regressionsmodelle für den Beispieldatensatz . . . . .	41
5.2.1.	Lineare Modelle . . . . .	42
5.2.2.	Geostatistik (räumliche Modelle pro Zeitpunkt) . . . . .	46
5.2.3.	additive Modelle . . . . .	48
5.2.4.	Vergleich der verwendeten Modellierungsstrategien . . . . .	50
5.2.5.	Auswertung der Modelle für alle Variablen . . . . .	51
5.3.	Anwendungsgebiet Regressionsmodelle . . . . .	55
<b>6.</b>	<b>Weiterführende Analysemethoden</b>	<b>57</b>
6.1.	Weitere Möglichkeiten zur Multiplen-Testen-Korrektur . . . . .	57
6.2.	Bayes und weitere Regressionsmodelle . . . . .	57
6.2.1.	Weitere Arten von Regressionsmodellen . . . . .	57
6.2.2.	Bayesianische Methoden . . . . .	58
6.3.	Weitere Methoden zur Beschreibung eines Unterschieds . . . . .	59
<b>7.</b>	<b>Zentrale Aussagen und Methodenüberblick</b>	<b>60</b>
7.1.	Vergleich und Beurteilung der Methoden . . . . .	60
7.1.1.	Ressourcenverbrauch . . . . .	60
7.1.2.	Modellierung von Korrelation in den Daten . . . . .	60
7.1.3.	Signifikanz bei niedrigem Signal-zu-Rausch-Verhältnis . . . . .	60
7.1.4.	Bewertung . . . . .	61
7.2.	Leitfaden für die Analyse . . . . .	63
7.3.	Zusammenfassende Auswertung des Beispieldatensatzes . . . . .	63
<b>A.</b>	<b>Weitere Abbildungen</b>	<b>65</b>
<b>B.</b>	<b>Abkürzungsverzeichnis und Liste der verwendeten R-Pakete</b>	<b>85</b>
<b>C.</b>	<b>Elektronischer Anhang</b>	<b>87</b>
<b>D.</b>	<b>Literaturverzeichnis</b>	<b>95</b>
<b>E.</b>	<b>Danksagung und Eigenständigkeitserklärung</b>	<b>100</b>

# 1. Einführung

## 1.1. Situationsbeschreibung

Die Atmosphäre eines Planeten ist ein chaotisches System, dessen Verhalten von zahlreichen interaktiv verknüpften Prozessen auf verschiedenen Skalen bestimmt wird, zum Beispiel Absorption von Sonnenstrahlung (Licht mit einer Wellenlänge von  $\approx 10^{-9}m$ ) oder Tiefdruckgebiete (Dynamik planetarer Wellensysteme mit einer Wellenlänge von  $\approx 10^6m$ ). Man kann mit der realen Atmosphäre allerdings keine physikalischen Experimente in dem Sinne durchführen, als dass man einen Parameter verändert und beobachtet, wie das System unter ansonsten unveränderten Umweltbedingungen darauf reagiert. Daher sind in der Klimaforschung computergestützte numerische Simulationen (Klimamodelle) von umso größerer Bedeutung.

Der Begriff chaotisches System wird hier in der mathematischen Bedeutung verwendet, dass eine kleine Änderung des Anfangszustandes zu völlig unterschiedlichen Endzuständen führen kann. Deswegen und aufgrund der Beschränkungen durch Rechenzeit, Speicherplatz und numerischer Genauigkeit kann das Verhalten der Atmosphäre mit dem Computer allerdings nur annähernd reproduziert werden.<sup>1</sup> Auch wenn in den letzten Jahrzehnten die Computerleistung rapide gestiegen ist und immer größere, physikalisch gesehen genauere und höher aufgelöste Modellsimulationen produziert werden, ist deshalb eine statistische Auswertung der Daten unabdingbar. Herausforderungen dabei sind nicht nur (inzwischen) die Größe der Datensätze, sondern vor allem die räumliche und zeitliche Abhängigkeiten innerhalb der Daten sowie die Unterscheidung zwischen natürlicher Klimavariabilität (Rauschen) und Klimaresponse (Signal).<sup>2</sup>

Diese Arbeit soll einen Überblick über bestehende statistische Verfahren geben, anschließend in einer Beispielstudie die Stärken und Schwächen dreier ausgewählter Verfahren aufzeigen und letztlich mit einem kurzen Leitfaden zur Auswahl geeigneter Verfahren in Abhängigkeit der Daten und Fragestellungen dem Leser praktische Hilfestellung bei seiner eigenen Arbeit geben.

## 1.2. Motivation

Klimamodellsimulationen werden hauptsächlich zur möglichst realistischen Reproduktion des realen Klimazustandes sowie zur Untersuchung des Einflusses verschiedener interner und externer Faktoren auf das System Atmosphäre verwendet. Zwei aktuelle Beispiele sind die Fragen, welche Auswirkungen eine Verdoppelung der  $\text{CO}_2$ -Emissionen hat oder ob die anthropogenen  $\text{CO}_2$ -Emissionen zu einer Veränderung des Klimas geführt haben können. Der in dieser Arbeit verwendete Datensatz fokussiert sich auf die Konzentrationsänderung von Spurengasen und Aerosolen, wie  $\text{SO}_x$ ,  $\text{NO}_x$  und Ruß (Black Carbon). Gemeinsam ist den Fragestellungen, dass zur Beantwortung die Ergebnisse zweier (oder mehrerer) Simulationen mit unterschiedlichem Input verglichen werden, in obigem ersteren Beispiel also die Ergeb-

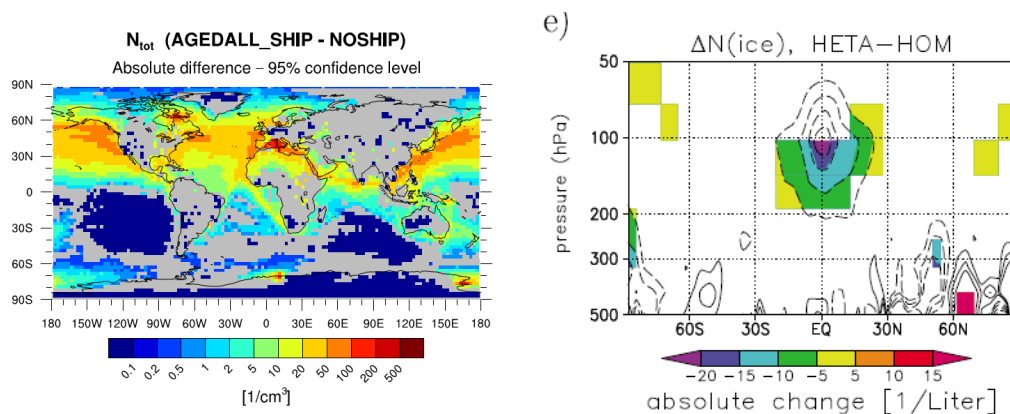
---

<sup>1</sup>Nähere Erläuterungen in 2.1

<sup>2</sup>Nähere Erläuterungen in 1.3

nisse einer zweiten Simulation mit doppelt so viel  $\text{CO}_2$  in der Atmosphäre als in der ersten Simulation. Zusätzlich zu der Frage, ob es einen Unterschied zwischen den Simulationen gibt, kann auch die Art des Unterschieds, zum Beispiel die räumliche Verteilung, von Interesse sein. Die Veränderung des Klimas kann dabei einerseits durch die Interpretation der direkten Veränderung der Simulationsergebnisse beschrieben werden, andererseits werden die Ergebnisse auch weiterverwendet, um mit einem Strahlungstransportmodell die Änderung im Energiehaushalt der Atmosphäre zu quantifizieren.

Üblicherweise wird der Unterschied zweier Simulationen als Differenz des Mittelwerts zweier Datenstichproben definiert (zum Beispiel die Differenz von 10 Jahresmittelwerten zweier Simulationen). Häufig werden die Mittelwerte pixelweise, also pro Gitterbox der Modelle, voneinander abgezogen und mit einem Student'schen t-Test auf statistische Signifikanz (in der Differenz) geprüft. Dabei stellte sich heraus, dass, vor allem bei geringen Unterschieden zwischen den Simulationsergebnissen und hoher natürlicher Variabilität innerhalb der jeweiligen Simulationen, nur sehr wenige Pixel als signifikant erkannt wurden und/oder die räumliche Verteilung physikalisch nicht sinnvoll interpretierbar war. Zwei Beispiele zur Illustration finden sich in Abbildung 1.1. Dies ist insofern für die weitere Verwendung der Simulationsergebnisse problematisch, da die Strahlungstransportmodelle zur Berechnung des Energiehaushalts räumlich vollständige Inputvariablen benötigen. Im Rahmen dieser Arbeit liegt der Fokus auf der Untersuchung zu Verbesserungsmöglichkeiten des t-Tests und zur Verwendbarkeit von alternativen Verfahren.



(a) Differenzen in der Anzahl von Aerosolpartikeln (in der untersten Atmosphärenschicht) (b) Differenzen in der Anzahl von Eiskristallen (zonal gemittelte Höhenverteilung)

Abb. 1.1.: (Univariater t-Test ( $\alpha = 0.05$  in 1.1a) und Student'scher t-Test ( $\alpha = 0.1$  in 1.1b) mit 10 Jahresmittelwerten. Vom jeweiligen Test als signifikant erkannte Pixel sind farbig gekennzeichnet. Abbildung 1.1a ist übernommen aus [53], Abbildung 1.1b aus [28]

### 1.3. Kriterien zur Methodenauswahl

Die Entscheidung für oder gegen die Verwendung eines bestimmten statistischen Verfahrens hängt hauptsächlich von den Eigenschaften der Daten, den verfügbaren

Computerressourcen und der zu bearbeitenden Fragestellung ab. Für jedes statistische Verfahren gibt es bestimmte Voraussetzungen an die Daten, die für seine korrekte Anwendung mehr oder weniger strikt erfüllt sein müssen. Viele Verfahren benötigen zum Beispiel normalverteilte Daten oder von einander unabhängige Beobachtungen. Hier werden nun fünf Herausforderungen, die normalerweise mit der Auswertung von Klimasimulationen verbunden sind, vorgestellt.

1. Eigenschaften der Daten:

a) Verteilung der Variablen:

Voraussetzung für viele gängige statistische Verfahren ist die annähernde Normalverteilung der Daten. Je nach Sensitivität bzw. Robustheit des Verfahrens dürfen die Daten mehr oder weniger davon abweichen. Bei vielen Methoden gibt es auch Erweiterungen für nicht-Normalverteilte Daten. Zählraten wie zum Beispiel Niederschlagsmenge oder Aerosolkonzentration sind im Allgemeinen nicht-Normalverteilt.

b) Korrelationen zwischen den Variablen:

Die meisten statistischen Verfahren beruhen auf der Annahme von i.i.d. (= independent identically distributed) Daten. Unabhängig bedeutet dabei, dass keine (oder nur geringe) Korrelationen zwischen den im Verfahren verwendeten Werten gibt. Dies stellt besonders für Klimamodellergebnisse eine Herausforderung dar, da diese Daten aus der Natur der Sache heraus meist sowohl eine räumliche als auch zeitliche Korrelation aufweisen.

2. Signal-zu-Rausch-Verhältnis:

Aufgrund der gestiegenen Computerleistung in den letzten Jahren werden immer komplexere Simulationen berechnet, mit zum Beispiel immer mehr verschiedenen Spurenelementen. Dabei wird dann der Effekt, hier die Konzentrationsänderung der Spurengase, mitunter so klein, dass er in der gleichen Größenordnung wie seine natürliche Variabilität oder sogar darunter liegt. Ein bekanntes und typisches Beispiel ist der Trend in der Temperatur, der durch eine steigende  $\text{CO}_2$ -Konzentration hervorgerufen wird. Wenn nur sehr wenig  $\text{CO}_2$  in der Simulation hinzugefügt wird, wird der Trend von den natürlichen Klimaschwankungen überlagert, und eine eindeutige Unterscheidung der beiden Simulationen ist schwer, wie man in den jeweiligen ersten Hälften der Graphen in Abbildung 1.2 sehen kann. Erst wenn sich die beiden Simulationen im  $\text{CO}_2$ -Gehalt deutlich unterscheiden (zweite Hälfte der Graphen), können sie eindeutig voneinander getrennt werden.

3. Computerressourcen:

Manche Verfahren benötigen mehr Computer- und Zeitressourcen als andere, wobei der Ressourcenverbrauch durchaus mit der Größe des Datensatzes skalieren kann. Dann steht die Zeit, die das Verfahren braucht, in Konkurrenz mit der, die man für die Auswertung aufwenden kann.

4. Definition der Fragestellung:

Es muss zu Beginn der Auswertung eindeutig festgelegt werden, wie „Unterschied“ in den Fragestellungen „Gibt es einen Unterschied?“ oder „Wie groß

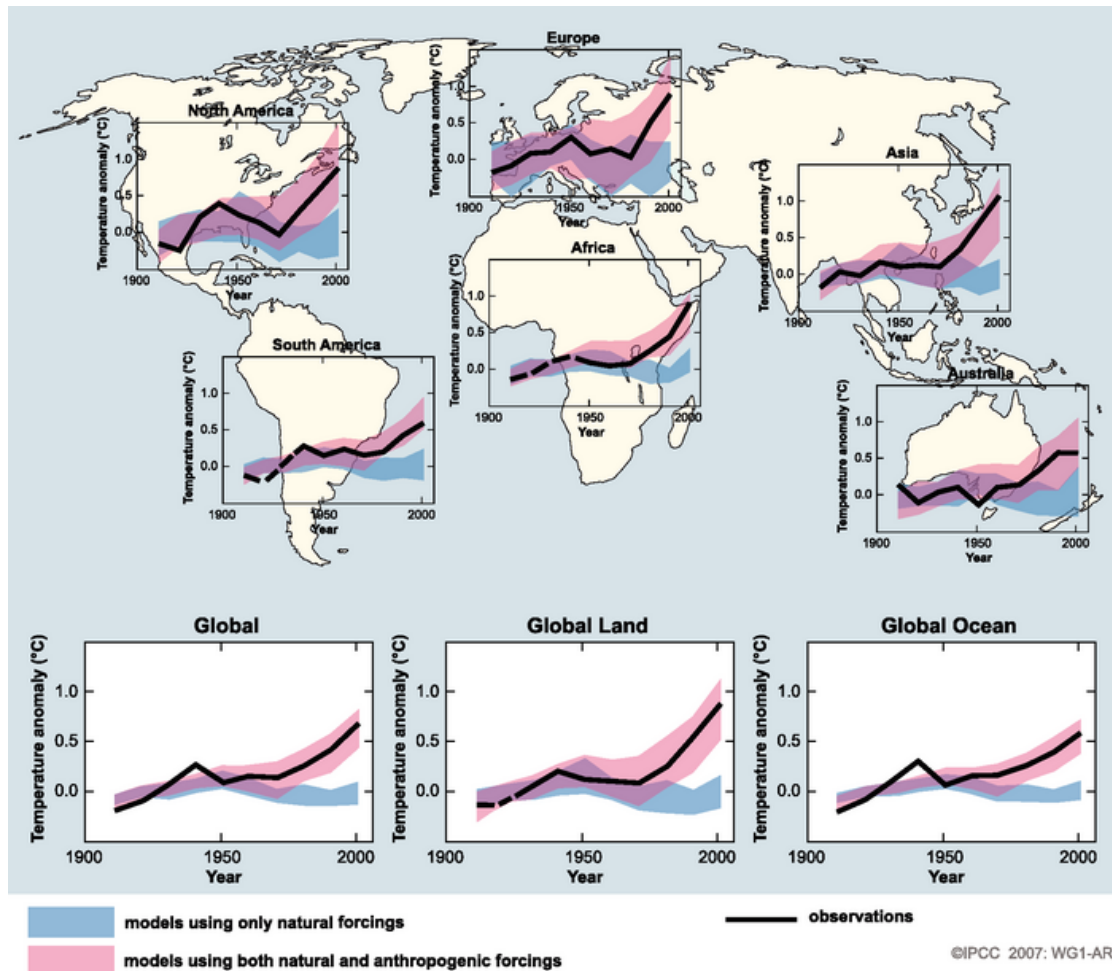


Abb. 1.2.: Temperature changes relative to the corresponding average for 1901-1950 ( $^{\circ}\text{C}$ ) from decade to decade from 1906 to 2005 over the Earth's continents, as well as the entire globe, global land area and the global ocean (lower graphs). The black line indicates observed temperature change, while the coloured bands show the combined range covered by 90% of recent model simulations. Red indicates simulations that include natural and human factors, while blue indicates simulations that include only natural factors. Dashed black lines indicate decades and continental regions for which there are substantially fewer observations. Grafik übernommen aus [27]

ist der Unterschied?“ definiert wird. Dabei gibt es verschiedene Möglichkeiten:

- a) Differenz im Mittelwert oder der Varianz (zum Beispiel beim Unterschied der Konzentration von Verkehrs-Emissionen)
- b) verschiedene Verteilungen, häufig von Extremwerten (zum Beispiel Niederschlagverteilung)
- c) unterschiedliche räumliche Muster durch Verschiebung, Ausdehnung oder Drehung (zum Beispiel räumliche Niederschlagsverteilung oder Wolkenbedeckung)
- d) unterschiedlicher zeitlicher Verlauf (zum Beispiel Verlauf der  $\text{CO}_2$ -Konzentration oder der globalen Mittel-Temperatur)

Hier ist schon zu sehen, dass bei der gleichen physikalischen Fragestellung wie zum Beispiel „Haben die Emissionen der Industrie einen Einfluss auf den Niederschlag?“, die Art des Einflusses und damit des Unterschieds zwischen den zwei Simulationen nicht eindeutig festgelegt ist. Bestenfalls wird diese Definition bereits in der Fragestellung eindeutig, wie bei den Formulierungen „Verschieben sich durch die Industrie-Emissionen die Niederschläge (räumlich)?“ oder „Wird durch die Industrie-Emissionen der Niederschlag intensiver?“.

Unter Berücksichtigung dieser Punkte wurden drei verschiedene statistische Verfahren zur Untersuchung ausgewählt, die im Folgenden kurz vorgestellt werden.

## 1.4. Verwendete Methoden

Der Student'sche t-Test ist ein sehr ressourcensparendes Verfahren, welches zur Differenz-Untersuchung verwendet werden kann. Allerdings beruht es auf Verteilungsannahmen, die bei Klimadaten meist verletzt sind. Auch für Daten mit geringem Signal-zu-Rausch-Verhältnis hat es sich, wie oben beschrieben, in der Praxis nicht bewährt. Zusätzlich muss beachtet werden, dass bei einer Simulation mit beispielsweise  $2^\circ$  räumlicher Auflösung 8100 t-Tests parallel durchgeführt werden, was dazu führt, dass das globale Signifikanzniveau nicht mehr mit dem eines einzelnen Tests übereinstimmt. Es werden im Laufe der Arbeit die Probleme des geringen Signal-zu-Rausch-Verhältnisses und des multiplen Testens betrachtet und Lösungsvorschläge untersucht.

Das Verfahren der Principal Components Analysis (PCA), in der Meteorologie auch als Empirical Orthogonal Functions (EOF) Analyse bekannt, beruht auf einer Projektion der Daten in einen Eigenraum und Reduktion auf wenige, aber viel Varianz erklärende, Komponenten. Diese werden dann mit einem multivariaten Test auf Signifikanz getestet. Damit kann die Größe des Datensatzes (und somit auch der Zeitaufwand für weitere Berechnungen) meist immens verringert werden. Hierbei ist ein geringes Signal-zu-Rausch-Verhältnis und eine nicht-Normalverteilung der Daten ein Problem. Es werden im Folgenden die Anwendbarkeit der PCA und eine Alternative, die Independent Components Analysis (ICA), untersucht.

Als zweite Alternative zum t-Test wird die Modellierung der Daten unter Verwendung von allgemeinen Regressionsmodellen untersucht. Die Modellierung mit

flexibel an die Daten anpassbaren Basisfunktionen (Generalised Additive Models, GAM), Gauß-Markov-Zufallsfeldern (Gaussian Markov Random Fields, GMRF) oder als Gauß'scher Prozess (Kriging) kann räumliche und zeitliche Korrelationen mit berücksichtigen. Es gilt jedoch, dass der Ressourcenaufwand (Zeit, CPUs und RAM-Speicher) mit der Komplexität des Modells steigt. Es wird untersucht, wie gut verschiedene Modellierungsansätze die Korrelation berücksichtigen können und wie groß der Ressourcenaufwand für die verschiedenen verwendeten Modelle und Algorithmen ist.

Am Ende der Arbeit findet sich ein Vergleich dieser drei Verfahren nach den obigen fünf Kriterien.

## 2. Klimamodellsimulationen

### 2.1. Einführung in Allgemeine Zirkulationsmodelle

Zur Simulation des Klimas werden numerische Modelle für die Atmosphäre verwendet, sogenannte Allgemeine Zirkulationsmodelle (General Circulation Models). Sie beruhen auf den grundlegenden Bewegungsgleichungen der Atmosphäre (Navier-Stokes-Gleichungen mit sphärischer Geometrie) und binden auch die wesentlichen thermodynamischen Prozesse (Strahlung, Phasenübergänge, Wärmeaustausch) mit ein. Dabei kann das Modell sowohl Rückkopplungs-Mechanismen für Wechselwirkungen, zum Beispiel Wasserdampf – Erwärmung, Wolken – Strahlung, Schnee- und Eisalbedo – Erwärmung, Strahlung – Chemie, mit dem Ozean, der Landoberfläche oder der Biosphäre enthalten, als auch an entsprechende Ozean-, Chemie-, oder Landnutzungsmodelle gekoppelt werden. Eine ausführliche Einführung findet sich beispielsweise in Güss et al. (1999) [24], woraus auch die Informationen in diesem Abschnitt stammen.

Der Zustand der Atmosphäre wird durch mehrere Variablen, wie beispielsweise Druck, Temperatur, Windgeschwindigkeit, Wasserdampfgehalt und Wolkenbedeckungsgrad beschrieben. Die Berechnung findet auf einem dreidimensionalen Gitter auf der Kugeloberfläche statt, typischerweise mit einer Auflösung von  $1^\circ$  bis  $5^\circ$  in Längen-/Breitengraden und 20 – 80 Schichten im Vertikalen. Dabei gibt es verschiedene Möglichkeiten zur Diskretisierung der Gittergeometrie (rechteckig, ikosaedrisch, variable Auflösung) und der Differentialgleichungen (Finite-Differenzen, spektral). Da die gewählte Auflösung eine Reihe kleinskaliger Prozesse, zum Beispiel Wolkenbildung, Diffusion oder Absorption von Strahlung nicht beschreiben kann, werden diese parametrisiert ins Modell mit aufgenommen.

Zusätzlich zu der aus diesen Approximationen durch die Parametrisierung, die Diskretisierung, die numerischen Lösungsverfahren und die vereinfachende Natur der Differentialgleichungen folgenden systemischen Unsicherheit verursacht die dem Problem inhärente chaotische Verhaltensweise der Atmosphäre eine statistische Unsicherheit in der Interpretation der Modellergebnisse.

Klimasimulationen können einerseits einen Gleichgewichtscharakter aufweisen, das heißt, sie verhalten sich stationär bei konstant bleibenden Rahmenbedingungen und zeigen somit eine stochastische Schwankung um den langzeitlichen Mittelwert. Andererseits können die Rahmenbedingungen auch eine zeitliche Veränderung aufweisen, wie zum Beispiel eine langsam steigende  $\text{CO}_2$ -Konzentration. Dann besitzt die Simulation einen transienten Charakter, der sich in der Reaktion des Modells auf die veränderten Rahmenbedingungen verursachten Trend der Klimavariablen zeigt. Es können auch die Zustandsvariablen in jedem Zeitschritt an reale Beobachtungsdaten angepasst werden. Dieses Vorgehen heißt „nudging“.

### 2.2. Beschreibung des Beispieldatensatzes

Der in dieser Arbeit verwendete Datensatz stammt aus einer Simulation von Righi et al. (2013) [53] mit der EMAC-MADE-Erweiterung des Klimamodell ECHAM5.



Es wurden zwei quasi-stationäre Simulationen unter Verwendung des Nudgings berechnet, wobei die Konzentrationen von strahlungswirksamen Spurgasen verändert wurden. Diese Konzentrationsänderung wirkt sich über die Veränderung der Energiebilanz auf das Verhalten des Systems Atmosphäre aus.

In einer Simulation sind zusätzlich zu den natürlichen und industriellen Emissionen auf der Landoberfläche die Emissionen des Schiffsverkehrs (über der Meeresoberfläche) als Input enthalten. Die sich daraus ergebende allgemeine physikalische Fragestellung „Haben die Schiffsverkehrsemissionen einen Einfluss auf das Klima?“ wird im Rahmen dieser Arbeit in die mathematische Frage „Ist die Differenz der beiden Simulationen signifikant verschieden von Null?“ umformuliert. Bei der Auswertung werden vor allem Unterschiede in der Konzentration von Sulfat-Aerosolen ( $\text{SO}_4$ ), Nitrat-Aerosolen ( $\text{NO}_3$ ), Ruß (Black Carbon, BC) und dem Wolkenbedeckungsgrad untersucht. Dabei dienen beim Vergleich der statistischen Verfahren die Sulfat-Konzentrationen als Beispiel für einen Datensatz mit hohem Signal-zu-Rausch-Verhältnis, also einem deutlich erkennbaren Unterschied zwischen den zwei Simulationsläufen, und der Wolkenbedeckungsgrad umgekehrt als Beispiel für ein geringes Signal-zu-Rausch-Verhältnis.

### 2.2.1. Beschreibung der Datenstruktur

Die hier ausgewerteten Simulationsergebnisse wurde auf einem globalem Gitter mit  $2,8125^\circ$  Auflösung in Breitengraden und  $2,7893^\circ$  in Längengraden in 19 Höhenschichten über den Zeitraum 1996 bis 2005 als monatliche Mittelwerte zusammengefasst. Daraus ergibt sich ein Datensatz mit  $128 \cdot 64 \cdot 19 \cdot 10 \cdot 12 = 18.677.760$  Einträgen pro Simulation. Die Höhenschichten sind in Bodennähe an die Orographie angepasst, in höheren Schichten nur an den atmosphärischen Druck, wie in Abbildung 2.1 dargestellt. Eine ungefähre Zuordnung der Höhenschichten zum vertikalen Aufbau der Atmosphäre findet sich in Tabelle 2.1.

Modelllevel	Druck (in hPa)	Höhe
19 - 16	1013 - 893	Planetare Grenzschicht (Boden bis ca. 1-2km): starke bodennahe Turbulenz
15 - 9/7/5	893 - 286/219/119	freie Troposphäre (zwischen Grenzschicht und Tropopause): gute Durchmischung, Bildung konvektiver Wolken, Abkühlung mit steigender Höhe
9/7/5 - 1	286/219/119 - 0	Stratosphäre (9km an den Polen/12km in den mittlere Breiten/17km in den Tropen bis ca. 40km): stabile Schichtung, kaum Durchmischung

Tabelle 2.1.: Approximative Zuordnung der Höhenlevel der Simulation zu den atmosphärischen Schichten, übernommen aus ??

Die  $\text{SO}_4$ -,  $\text{NO}_3$ - und BC-Konzentrationen sind als Massenmischungsverhältnis in  $\frac{\text{mol}}{\text{mol}}$  gegeben, können bei Bedarf aber auch unter Verwendung des Druckes  $p$  und der Temperatur  $T$  mit der Formel  $c[\frac{\mu\text{g}}{\text{m}^3}] = \frac{M_c \cdot p}{R \cdot T} \cdot c[\frac{\text{mol}}{\text{mol}}]$  in Dichten umgerechnet werden, wobei  $M_c$  die Atommasse des jeweiligen Moleküls (für Ruß  $470 \frac{\text{g}}{\text{mol}}$ ) und  $R =$

$8.314459848 \frac{J}{mol \cdot K}$  die allgemeine Gaskonstante ist. Der Wolkenbedeckungsgrad ist in Bruchteilen von 1 gegeben.

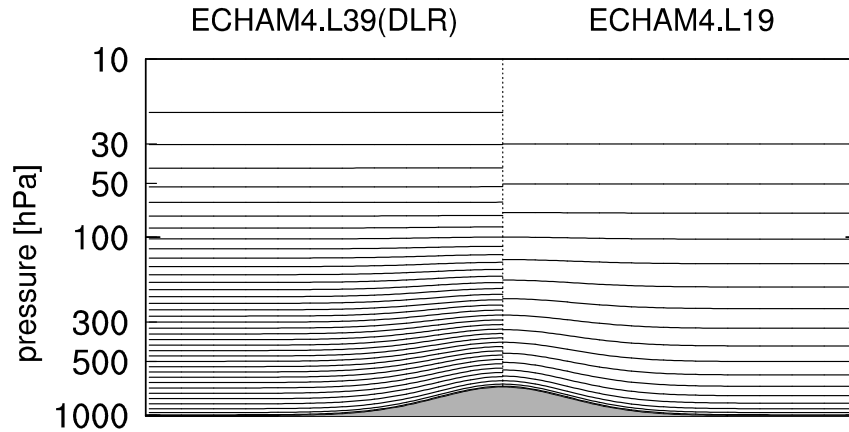


Abb. 2.1.: Höhenlevel des verwendeten Klimamodells (rechte Seite) mit idealisierter Orographie, übernommen aus [37]

Zusätzlich wurde aus der Liste der verfügbaren Modellvariablen die  $SO_2$ -Konzentration als approximative räumliche Verteilung der Schiffsverkehrsemissionen einerseits und der natürlichen und industriellen Quellen über Landflächen andererseits verwendet. Als Annäherung an den Prozess der Auswaschung von Partikeln aus der Atmosphäre, welcher die wichtigste physikalische Senke für Aerosole ist, wurde die großskalige Niederschlagsverteilung (Summe von Regen und Schnee, in  $\frac{kg}{m^2 s}$ ) verwendet. Da die Emissionskonzentration und Niederschlagsverteilung nur als Proxy für die Quellen und Senken dienen, wurden sie vor der Verwendung in den Regressionsmodellen gemäß der Formel  $Z = \frac{X - \hat{\mu}}{\hat{\sigma}}$  standardisiert. Dabei ist  $\hat{\mu}$  der Mittelwert und  $\hat{\sigma}$  die Standardabweichung der Daten X. Eine typische räumliche Verteilung der drei standardisierten Variablen ist in Abbildung 2.2 dargestellt. Weil das Interesse an der räumlichen und zeitlichen Verteilung und nicht am Absolutwert besteht, führt dies zu besserer Vergleichbarkeit der Koeffizienten im Regressionsmodell und teilweise auch zu höherer numerischer Stabilität der Modelle (sh. 6.2.2). Alternativ wurde auch eine Einteilung in vier Stufen mit den 50%-, 75% und 95%-Quantilen beim Schiffsverkehr und 50%-, 90%- und 99%-Quantilen bei den Landemissionen und Niederschlag als Schnittpunkte und eine Diskretisierung durch Runden der standardisierten Variablen auf ganze Zahlen verwendet. Außerdem wurden die standardisierte Emissionsmuster mit jeweils drei verschiedenen Gauß- und Mittelwertsfiltern geglättet<sup>1</sup> (Tabelle 2.3). Es ging dabei die Überlegung mit ein, nicht nur den direkten Einfluss der Emissionen des jeweiligen Pixels quantifizieren zu wollen, sondern durch die räumliche Glättung des Emissionsmusters auch den der Nachbapixel. Der Unterschied zwischen den Filtern ist die Anzahl und Gewichtung der Nachbapixel. Dabei ist sowohl die Wahl der Schnittpunkte bei der Diskretisierung als auch der Größe und Standardabweichung der Gauß-Filter völlig willkürlich. Einen Vergleich der räumlichen Glättung liefert die Abbildung A.1.

<sup>1</sup>Ein Teil des Algorithmus stammt von

<http://lists.r-forge.r-project.org/pipermail/rcpp-devel/2012-January/003312.html>

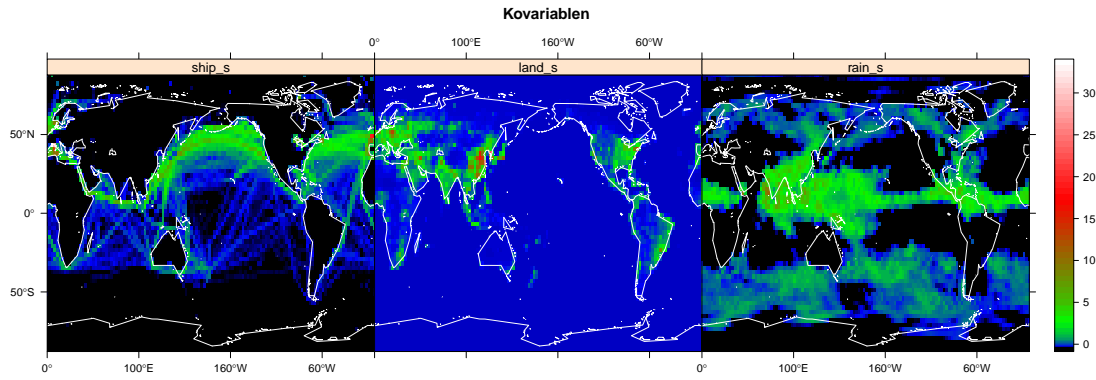


Abb. 2.2.: Räumliche Verteilung der standardisierten Schiffs- und Landemissionen und des Niederschlags für Juli. Den gesamten Jahresgang sieht man in C.1.1.

Gauß-Filter			Mittelwerts-Filter mit nächsten Nachbarn		
Name	Größe in Pixel	$\sigma$	k1	k2	k3
small	3	1	$\frac{1}{4} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	$\frac{1}{4} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$	$\frac{1}{16} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$
medium	5	1			
big	7	3			

Abb. 2.3.: Verwendete Gauß- und Mittelwertsfilter

## 2.2.2. Eigenschaften der Variablen

### 2.2.2.1. Verteilung der Aerosol- und Wolken-Variablen

Keine der untersuchten Variablen besitzt eine empirische Verteilung, die durch eine Normalverteilung beschrieben werden kann. Eine grafische Darstellung der Histogramme findet sich in C.1.2. Da negative Konzentrationen und Bedeckungsgrade physikalisch nicht möglich sind, ist der Wertebereich  $[0; +\infty)$ , innerhalb dessen eine starke Linksschiefe der Verteilung zu sehen ist. Logarithmiert man die Variablen, wird die Verteilung symmetrischer, ist aber spitzer als eine Normalverteilung. Auch die Verteilungen Differenz der Variablen, obwohl nun im Wertebereich  $(-\infty; +\infty)$ , weisen eine Linksschiefe und größere Steilheit als eine Normalverteilung auf.

Bei Betrachtung der Verteilung der Differenz in einzelnen Pixeln zeigt sich ein ähnliches Bild. Ein Shapiro-Wilk-Test auf Normalverteilung der Jahresmittelwerte in der untersten Höschicht lehnt die Normalverteilungsannahme zum Signifikanzniveau  $\alpha = 0.05$  auch mit Korrektur für multiples Testen nur bei wenigen Pixeln, meist an den Polen oder in Äquatornähe, ab, wie in Abbildung A.2 zu sehen ist. Dies liegt vermutlich an der Stichprobengröße von  $n = 10$ , da ein Shapiro-Wilk-Test bei allen Monaten ( $n = 120$ ) bei circa 80% aller Pixel die Normalverteilung ablehnt (vgl. Abbildung A.3). Allerdings ist bei der größeren Stichprobe die zeitliche Korrelation nicht mit berücksichtigt. Diese beiden Abbildungen beschreiben das Testergebnis für die Sulfat-Aerosole, die Ergebnisse für die anderen drei Variablen finden sich in Anhang C.2.1.

Auf jeden Fall sollte die etwaige nicht-Normalverteilung bei der Anwendung der statistischen Verfahren und der Interpretation der Ergebnisse berücksichtigt werden.

#### **2.2.2.2. Räumliches und zeitliches Verhalten der Aerosol- und Wolken-Variablen**

Im globalen Mittel betrachtet, sinkt der Wert der Konzentrationen mit steigender Höhe, was auch bedeutet, dass die absolute Differenz geringer wird. Zusätzlich sinkt aber auch die relative Differenz. Besonders gut sieht man dies bei den Sulfat-Aerosolen in Abbildung A.4. Einzig die Nitrat-Konzentrationen steigen in der oberen Atmosphäre wieder an. Die Differenz des Wolkenbedeckungsgrades variiert im Mittel am schwächsten mit der Höhe. Grafisch dargestellt ist dies in C.1.3.

Betrachtet man das räumliche Muster der pixelweisen Jahresmittel der Konzentrationen, so zeigt sich, unabhängig von der betrachteten Höschicht, keine große Variation zwischen den Jahren. Bei den pixelweisen Mittelwerten der jeweiligen Monate hingegen lässt sich ein klarer Jahresgang erkennen. Dieses Verhalten zeigen auch die Differenzen. Lediglich beim Wolkenbedeckungsgrad kann der Jahresgang der Variablen an sich noch einigermaßen gut erkannt werden, bei der Differenz ist jedoch kein solcher mehr zu sehen. Eine ausführliche grafische Darstellung findet sich in C.1.4, im Folgenden wird kurz die Verteilung der Sulfat-Konzentration, wie sie in den Abbildungen 2.4, A.6 und A.5 zu sehen ist, beschrieben.

Räumliche Maxima der Differenz sind in den unteren Höschichten hauptsächlich über den großen Industriegebieten über Land, wie zum Beispiel Europa und China, zu finden. Diese ziehen sich mit größer werdendem Radius bis in die mittlere Troposphäre hinein. Nur in den unteren Schichten findet man auch über den Ozeanen, vor allem dem Nordpazifik, Nordatlantik und Mittelmeer, und in Küstennähe von Indien, Südamerika und Afrika, ausgedehnte Gebiete mit relativ hohen Differenzen. Diese entsprechen in etwa dem Muster der Schiffsemissionen. Je weiter man sich vom Boden entfernt, desto räumlich homogener werden die Differenzmuster, besonders in zonaler Richtung. Der Jahresgang ist in allen Höschichten klar ausgeprägt. In den unteren Schichten schwankt vor allem die Stärke der Differenz über den Ozeanen, während in den höheren Schichten ein Wechsel des Maximas der Differenz zwischen Äquator und Nordhalbkugel zu sehen ist.

Zum Vergleich zeigt Abbildung A.7 die Monatsmittel des Wolkenbedeckungsgrades in der zweit-untersten Höschicht. Hier ist kein klarer Jahresgang mehr zu erkennen, obwohl die aufeinander folgenden Monate leichte Ähnlichkeit aufweisen. Im Gegensatz zu den großskaligen räumlichen Strukturen der Konzentrationsdifferenzen ist die Differenz des Wolkenbedeckungsgrades räumlich sehr variabel.

#### **2.2.2.3. Räumliche und zeitliche Korrelation innerhalb der Aerosol- und Wolken-Variablen**

**Zeitliche Korrelation** Der schon in der Darstellung der des mittleren Jahresverlaufs der Konzentrationen auffällige Jahresgang zeigt sich in der empirischen Autokorrelationsfunktion (ACF) der Variablen im globalen Mittel sehr deutlich. Bei der Differenz ist er bei  $\text{NO}_3$  schwächer und beim Wolkenbedeckungsgrad in der ACF nicht mehr zu sehen. Zusätzlich ist, außer bei der Differenz des Wolkenbedeckungsgrades, eine zeitliche Korrelation von bis zu zwei Monaten vorhanden. Als

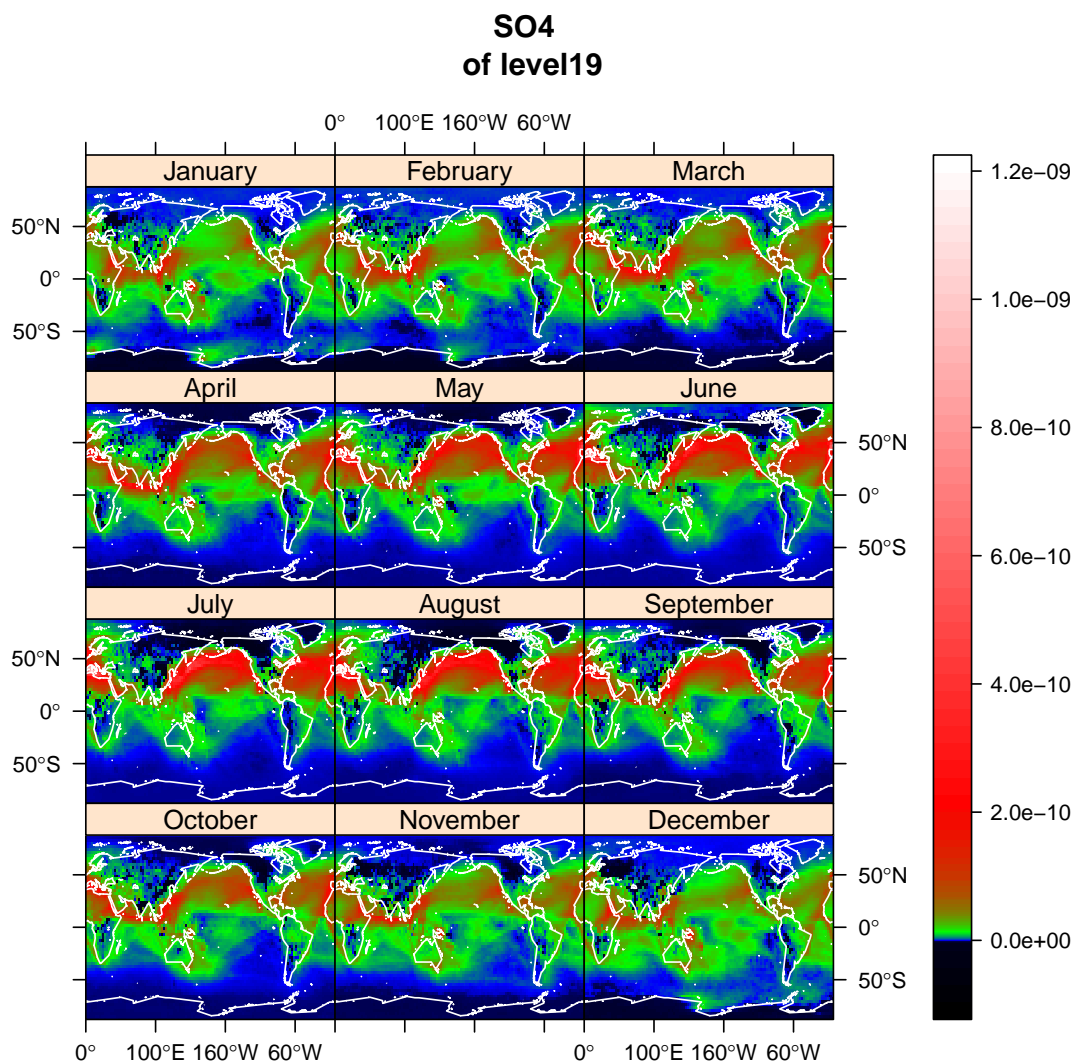


Abb. 2.4.: Räumliche Verteilung der Differenz der Sulfat-Konzentration in [mol/mol], gemittelt über die jeweiligen Monate in der untersten Höhengschicht

Beispiel ist die ACF für die unterste Höhengschicht in Abbildung A.8 aufgetragen, eine komplette Grafik findet sich in C.1.5.

**Korrelation in z-Richtung (vertikal)** In Übereinstimmung mit der physikalischen Aufteilung der Atmosphäre in die verschiedenen Höhenbereiche (Tabelle 2.1) zeigen auch die räumlichen Verteilungen der einzelnen Variablen und ihrer Differenzen pro Modellschicht innerhalb dieser physikalischen Höhenbereiche große Ähnlichkeit und zwischen den Höhenbereichen weitgehende Unähnlichkeit. Besonders die räumliche Verteilung der unteren drei bis vier Modell-Höhengschichten ist sich sehr ähnlich. Definiert wurde hierbei Ähnlichkeit als nahe beieinander liegende Werte der Pearson-Korrelation  $r^2$  und einem Modified Index of Agreement  $d \approx 1$ .

Der Modified Index of Agreement ist als  $md = \frac{1 - (\sum_j (obs1 - obs2)^j)}{\sum_j (|obs2 - \text{mean}(obs1)| + |obs1 - \text{mean}(obs1)|)^j}$  für zwei Datenvektoren obs1 und obs2 definiert [70]. Die Tabellen für beide Maßzahlen finden sich in C.1.6.

Aufgrund der Unterschiedlichkeit der Höhenbereiche kann die Korrelation der Daten in z-Richtung (vertikal) in der Analyse berücksichtigt werden, in dem für jeden Bereich eine Schicht exemplarisch analysiert wird oder auch durch Mittelung der jeweiligen Schichten innerhalb der einzelnen Höhenbereiche. Im Rahmen dieser Arbeit werden die Analysen exemplarisch für die unterste Schicht beschrieben und jeweils mit den Ergebnissen der Analysen für die höheren Schichten verglichen.

**Globale Korrelation** Die globale räumliche Korrelation innerhalb einer Höhengschicht (für einen Zeitpunkt) kann auf zwei verschiedenen Weisen beschrieben werden. Einmal gibt es die Möglichkeit mit lokalen Indizes, wie zum Beispiel Moran's I und Geary's C, die Stärke der Korrelation zwischen einem Pixel und seinen Nachbarn zu messen [4]. Je extremer der resultierende lokale Index, desto größer ist der Einfluss der jeweiligen Region auf die Gesamtstatistik. Zur Illustration ist in Abbildung A.9 das lokale Moran-I und Geary-C der Sulfat-Differenz in der untersten Höhengschicht aufgetragen. Die restlichen Abbildungen sind in C.1.7 zu finden. Des weiteren kann das empirische Variogramm  $2\hat{\gamma}(h) \equiv \text{ave}((X(s_i) - X(s_j))^2 : ||s_i - s_j|| \leq h; h \text{ klein})$  der Daten  $\mathbf{X}$  an den Orten  $\mathbf{s}$  für verschiedene Abstände  $h$  berechnet werden [16].  $\hat{\gamma}(h_i)$  ist ein Maß der Punkt-zu-Punkt-Varianz für eine feste Entfernung  $h_i$ , und somit ein inverses Maß für die Ähnlichkeit zwischen den allen Punkten mit Abstand  $h_i$ . Das heißt, je größer der Wert  $\hat{\gamma}(h_i)$ , desto kleiner die räumliche Korrelation. Dabei ist es auch möglich,  $\hat{\gamma}$  in nur einer räumlichen Dimension zu berechnen, zum Beispiel meridional oder zonal, um zu untersuchen, ob die räumliche Korrelation isotrop ist. Bei der Interpretation ist zu beachten, dass für die gerichteten Berechnungen sowie große Abstände zwischen den betrachteten Punkten weniger Punktpaare zur Verfügung stehen und der berechnete Wert für  $\hat{\gamma}$  somit mit einer größeren Unsicherheit behaftet ist als der für kleine Abstände und omnidirektionale Berechnungen.

Tendenziell zeigen die Variogramme der Differenzen für die kurzen Abstände kaum anisotropes Verhalten, wie in Abbildung A.11 zu sehen ist. Lediglich für größere Entfernungen zeigen die Punkt-zu-Punkt-Varianz für die Ost-West-Richtung ein anderes Verhalten:  $\hat{\gamma}$  der Konzentrationsdifferenzen steigt dabei für die oberen Höhengschichten schneller an als für die unteren und während es für die Ost-West-Richtung meist weiter ansteigt, sinkt  $\hat{\gamma}$  für die anderen Richtungen ab einer Entfernung von ungefähr  $180^\circ$  wieder. Es muss bei der Interpretation jedoch bedacht werden, dass ab einer Entfernung von  $180^\circ$  weniger Punktpaare in Nord-Süd-Richtung zur Berechnung zur Verfügung stehen.  $\hat{\gamma}$  der Differenz des Wolkenbedeckungsgrades zeigt ein größtenteils konstantes Verhalten. Dies deckt sich mit der Beobachtung, dass die Differenz des Wolkenbedeckungsgrades eine kleinskalige räumliche Struktur aufweist, die der Konzentrationen eine großskalige und somit eine höhere Korrelation für näher beieinander liegende Werte. Die Variogramm-Abbildungen für alle Höhengschichten finden sich in C.1.7.

**Raum-Zeit-Korrelation** Die Methodik des empirischen Variogramms kann auch raum-zeitlich verwendet werden, wobei von Isotropie in der räumlichen Dimension ausgegangen wird. Dabei wird die Zeit mathematisch gesehen als eine weitere Raumdimension betrachtet. Von Interesse ist dabei, ob sich die Stärke der räumlichen Korrelation mit der Zeit verändert, also beispielsweise die Werte der Nitratkonzentration im Sommer über größere Entfernungen miteinander korrelieren als

im Winter. Dabei zeigte sich, dass in den unteren Höhenschichten eine solche Variation nicht vorhanden ist, in den oberen jedoch schon (Grafiken C.1.8). Dabei zeigte die Differenz eine größere zeitliche Variabilität der räumlichen Korrelation als die Variablen an sich, vor allem in den oberen Schichten (Abbildung A.11). In den unteren Schichten können die räumliche und die zeitliche Korrelation also getrennt modelliert werden, wohingegen bei höheren Schichten eine raum-zeitlich veränderbare Komponente mit berücksichtigt werden muss.

#### **2.2.2.4. Korrelationen zwischen den Emissions- und Niederschlagsmustern**

Eine Voraussetzung für die Verwendung von Größen als Kovariablen in (linearen) Regressionsmodellen ist ihre gegenseitige Unkorreliertheit, da sonst numerische Instabilitäten bei der Berechnung des Modells und Identifikationsprobleme bei der Interpretation auftreten können. Die räumlichen Muster von Land- und Schiffsemissionen sind so gut wie nicht korreliert, was sich aus ihrer geographischen Aufteilung in Land- und Ozeanoberfläche ergibt. Die Niederschlagsmuster sind mit den Emissionsmustern je nach Monat zwischen 0% und 20% korreliert (Pearson-Koeffizient). Dass die geglätteten Schiffsmuster untereinander hoch korreliert (zwischen 80% und 90%) sind, erklärt sich durch das Verfahren der Glättung.

#### **2.2.3. Wichtige Merkmale des Datensatzes**

Aus obiger deskriptiver Analyse haben sich drei wichtige Merkmale des Beispieldatensatzes herausgehoben, welche im Laufe weiterer Analysen nicht unberücksichtigt bleiben dürfen. Es zeigte sich, dass neben eindeutigen zeitlichen Mustern wie dem Jahresgang auch eine räumliche Korrelation besteht. Diese ist allerdings je nach betrachteter Höhenschicht und Variable unterschiedlich. Des weiteren sind die Daten nicht gut durch eine Normalverteilung zu beschreiben, und zeigen auch wieder je nach Höhenschicht und Variablen einen unterschiedlich großen Wertebereich mit teilweise sehr extremen Ausreißern (Grafik C.1.3). Zusätzlich ist die Differenz des öfteren sehr klein im Vergleich zu den Werten an sich (Tabelle C.1.3) und der Schwankungsbreite der Daten, das heißt die Daten haben teils ein sehr kleines Signal-zu-Rausch-Verhältnis. Dies sind die ausschlaggebendsten Herausforderungen, welche bei der jetzt folgenden Anwendung der vorgeschlagenen statistischen Verfahren berücksichtigt werden müssen.

# 3. Hypothesentests

## 3.1. Mathematischer Hintergrund

### 3.1.1. Allgemeine Theorie statistischer Hypothesentests

Die allgemeine Theorie statistischer Hypothesentests und auch spezifische Testprozeduren können ausführlicher in Lehrbüchern zur induktiven Statistik nachgeschlagen werden; die folgende Ausführung ist verkürzt übernommen aus Toutenburg et al. (2008) [67]. Ein Hypothesentest (hier: Signifikanztest) dient zur Entscheidungsfindung in Bezug auf eine mathematisch formulierte Hypothese, die Nullhypothese  $H_0 : \theta \in \Theta_0$ , und ihre Alternative  $H_1 : \theta \in \Theta_1$ , wobei sich  $\Theta_0$  und  $\Theta_1$  gegenseitig ausschließen, aber den ganzen Wertebereich  $\Theta$  überspannen. Aus den gemessenen Daten  $x$ , einer Stichprobe aller möglichen Datenwerte  $X$ , wird eine Prüfgröße  $t(x)$  als Realisierung der Teststatistik  $T(X)$  berechnet.  $T(X)$  ist eine Zufallsvariable, die unter Annahme der Richtigkeit der Nullhypothese einer festgelegten Verteilung  $F(T)$  folgt. Die Art der Berechnung von  $T(X)$  und Typ der Verteilung  $F(T)$  hängt von der jeweiligen Testprozedur ab und bei parametrischen Testprozeduren zusätzlich von der angenommenen Verteilung von  $X$ . Nicht-parametrische Testprozeduren müssen verwendet werden, wenn die Verteilung der Daten nicht bekannt oder aus  $x$  zu bestimmen ist, oder wenn für sie keine parametrische Testprozedur existiert.

Die Testentscheidung wird aus dem Vergleich der berechneten Prüfgröße  $t$  mit dem theoretischen Wertebereich von  $T$ , welcher in einen Annahme- und Ablehnungsbereich geteilt wird, getroffen. Der kritische Wert, welcher Annahme- und Ablehnungsbereich voneinander trennt, wird über die Einschränkung der Wahrscheinlichkeit, einen Fehler 1. Art zu begehen, festgelegt. Diese Schranke ist das Signifikanzniveau  $\alpha$  des Tests. Ein Fehler 1. Art ist als diejenige Fehlentscheidung definiert, die  $H_0$  ablehnt, obwohl  $H_0$  eigentlich zutreffen würde (Tabelle 3.1). Die Wahrscheinlichkeit  $\beta$ , einen Fehler 2. Art zu begehen, ist dabei unter anderem invers abhängig von  $\alpha$ . Die Größe  $1 - \beta$ , auch Power, Sensitivität oder Teststärke genannt, welche die Wahrscheinlichkeit für eine 'True Negative'-Entscheidung angibt, wird zum Vergleich verschiedener Testprozeduren verwendet.

		Wahrheit	
		$H_0$ richtig	$H_0$ falsch
Testentscheidung	$H_0$ nicht abgelehnt	True Positive	False Negative (Fehler 2. Art)
	$H_0$ abgelehnt	False Positive (Fehler 1. Art)	True Negative

Tabelle 3.1.: Fehlentscheidungen beim Hypothesentesten

Ein alternatives Verfahren, die Testentscheidung zu treffen, welches meist in den gängigen Statistiksoftware-Programmen verwendet wird, ist der direkte Vergleich des aus  $t(x)$  berechneten p-Wertes mit dem Signifikanzniveau. Der p-Wert ist dabei die Wahrscheinlichkeit, unter Annahme der Nullhypothese einen genauso großen oder extremeren Wert als  $t$  als Prüfgröße zu erhalten. Die Definition von 'extrem'



hängt dabei von der Struktur der Hypothesen  $H_0$  und  $H_1$  ab. Man spricht von einem zweiseitigen Test, falls  $H_0 : \theta = \theta_0$  und  $H_1 : \theta \neq \theta_0$  die Situation beschreiben, dass  $\theta$  von einem vorher festgelegten Wert  $\theta_0$  abweicht. Ist die Abweichung einseitig, also  $\theta \geq \theta_0$  oder  $\theta \leq \theta_0$ , so heißt auch der zugehörige Test einseitig. Dabei gilt, dass ein einseitiger Test bei gleicher Prozedur und Stichprobe eine höhere Teststärke, also kleineres  $\beta$ , hat als ein zweiseitiger.

### 3.1.2. Multiples Testen

Beim simultanen Testen mehrerer Hypothesen mit der gleichen Prozedur unter Verwendung des gleichen Datensatzes muss zwischen den Signifikanzniveaus  $\alpha_{\text{einzel}}$  der jeweiligen Hypothesen und dem Signifikanzniveau  $\alpha_{\text{gesamt}}$  der übergreifenden Hypothese unterschieden werden. Die Situation ist in Tabelle 3.2 veranschaulicht. Unter der Annahme, dass alle  $H_0$  richtig sind, würde man beim Testen von  $m$  Hypothesen also  $V = \alpha_{\text{gesamt}} \cdot m$  abgelehnte  $H_0$  erwarten. Man möchte durch die Festlegung von einem  $\alpha_{\text{gesamt}}$  also die erwartete Anzahl  $E(V)$  von Fehlern 1. Art nach oben beschränken.

Dabei gibt es verschiedene Konzepte, um  $\alpha_{\text{gesamt}}$  zu definieren. Die Family-Wise Error Rate  $FWER = P(V \geq 1)$  ist die Wahrscheinlichkeit, mindestens einen Fehler 1. Art zu begehen und kommt damit dem Konzept des Signifikanzniveaus bei einzelnen Tests am Nächsten. Alternativ kann auch die False Discovery Rate  $FDR = E[\frac{V}{R}]$ , der erwartete Anteil von fälschlicherweise abgelehnten Nullhypothesen an allen abgelehnten Nullhypothesen, betrachtet werden. Im Vergleich zur Beschränkung der FWER führt die Beschränkung der FDR zu einer größeren Teststärke, aber auch mehr Fehlern 1. Art bei den Einzelhypothesen. Im Folgenden werden einige der häufig angewendeten Korrekturprozeduren für multiples Testen vorgestellt. Einen Überblick über die Thematik sowie die einzelnen Korrekturmethode gibt das Review-Paper von Shaffer (1995) [62], aus welchem auch die folgenden Erklärungen übernommen sind.

		Wahrheit		
Testentscheidung		$H_0$ richtig	$H_0$ falsch	Gesamt
	$H_0$ nicht abgelehnt	U (True Positive)	T (False Negative)	m-R
	$H_0$ abgelehnt	V (False Positive)	S (True Negative)	R
Gesamt		$m_0$	$m - m_0$	m

Tabelle 3.2.: Anzahl der Fehlentscheidungen bei multiplem Testen mehrerer Hypothesen

#### 3.1.2.1. FWER-Kontrollierende Korrekturen

Die Family-Wise Error Rate  $FWER = P(V \geq 1)$  wird, ausgehend von der Bonferroni-Ungleichung [62], durch  $\sum \alpha_{\text{einzel}}$  nach oben beschränkt. Die daraus folgenden Bonferroni-Korrektur  $\alpha_{\text{einzel}} = \frac{\alpha_{\text{gesamt}}}{m}$  reduziert jedoch die Teststärke der Einzeltests (über das inverse Verhältnis von  $\alpha$  und  $\beta$ ) übermäßig und wird daher nicht zur Verwendung empfohlen.

**Holm-Bonferroni Step-down Korrektur** Die Reduktion der Einzelteststärke kann durch eine sequentielle Vorgehen umgangen werden. Zuerst werden alle Einzelhypothesen aufsteigend nach ihren p-Werten sortiert und dann der Reihe nach mit einem Signifikanzniveau  $\alpha_{korrigiert}$  verglichen. Sobald der i-te p-Wert  $\alpha_{korrigiert}$  überschreitet, also  $p_i > \alpha_{korrigiert} = \frac{\alpha}{m-i+1}$ , werden alle Hypothesen mit p-Werten kleiner als oder gleich  $\alpha_{korrigiert}$  abgelehnt, also alle  $H_1, \dots, H_{i-1}$ . Wird die Schranke schon bei der ersten Hypothese überschritten, so werden alle Hypothesen nicht abgelehnt.

**Šidák-Korrektur** Sowohl die Bonferroni- als auch die Holm-Prozedur sind für beliebige Assoziationsstrukturen der Hypothesen anwendbar. Falls die Teststatistiken der Einzelhypothesen unabhängig oder positiv voneinander abhängig sind, können beide durch Ersetzen von  $\frac{\alpha}{m}$  mit  $1 - (1 - \alpha)^{1/m}$  leicht verbessert werden, wobei der Effekt für kleine  $\alpha$  nur unwesentlich ist.

**Hochberg-Korrektur** Die Hochberg-Korrektur ist ähnlich der Holm-Korrektur, nur dass hier im i-ten Schritt alle  $H_1$  bis  $H_i$  abgelehnt werden, falls die obige Stoppbedingung erreicht ist. Aufgrund des leicht veränderten Ablehnungsbereichs ist diese Prozedur keine abgeschlossene mehr und damit nur für unabhängige oder positiv assoziierte Teststatistiken gültig. Eine Alternative mit größerer Teststärke zur Hochberg-Korrektur ist die Hommel-Korrektur, für welche die gleichen Bedingungen gelten.

### 3.1.2.2. FDR-Kontrollierende Korrekturen

Im Gegensatz zu den FWER-Kontrollierenden Korrekturen erlauben die FDR-Methoden eine höhere Rate an Fehlern 1. Art bei den Einzelhypothesen, erlauben damit aber auch eine höhere einzelne Teststärke. Zurückgehend auf die ursprüngliche Methode von Benjamini und Hochberg [7] für unabhängige Hypothesen können verschiedene Modifikationen, zum Beispiel für abhängige Hypothesen [8], angewendet werden.

Ebenso wie bei der FWER-Korrektur von Hochberg werden bei der FDR-Korrektur die p-Werte der einzelnen Hypothesen sortiert und der Reihe nach mit einem  $\alpha_{korrigiert}$  verglichen.  $\alpha_{korrigiert}$  ist hier gleich  $\frac{i \cdot \alpha}{m \cdot c(m)}$  für die i-te Hypothese.  $c(m)$  hängt dabei von der Korrelationsstruktur der Hypothesen ab:

$$c(m) = \begin{cases} 1 & \text{bei positiver Assoziation oder Unabhängigkeit} \\ & \text{(ursprüngliche Benjamini-Hochberg-Korrektur)} \\ \sum_{i=1}^m \frac{1}{i} & \text{bei unbekannter Abhängigkeit} \\ & \text{(Benjamini-Hochberg-Yekutieli-Korrektur)} \end{cases}$$

Ausgefeiltere Korrekturen, zum Beispiel über Gauß-Markov-Zufallsfelder, welche die FDR kontrollieren, wurden im Zusammenhang mit der Auswertung von fMRI-Gehirnscans entwickelt [5]. Aus der Geostatistik kommt die Idee, die räumliche Struktur der Teststatistiken über eine Transformation in den Wavelet-Raum für eine Anwendung bei geringem Signal-zu-Rausch-Verhältnis auszunutzen [63]. Beide Methoden werden kurz in Kapitel 7 angesprochen.

### 3.1.3. Der Student'sche t-Test und Alternativen

Häufig steht die Bezeichnung t-Test synonym für verschiedene Arten, Hypothesen über den Lageparameter (Mittelwert oder Median) einer Stichprobe zu testen. Die verschiedenen Testprozeduren, welche unter dem Schlagwort t-Test zusammengefasst sind, können in einführender Literatur zur Statistik nachgeschlagen werden. Hier wurde Toutenburg et al. (2008) [67] verwendet. Es können entweder der Mittelwert einer Stichprobe  $X$  mit einem theoretischen Wert  $\mu_0$  oder die Mittelwerte zweier Stichproben  $X$  und  $Y$  miteinander verglichen werden. Im Fall zweier Stichproben wird zwischen der Situation gleicher oder unterschiedlicher Varianz der Stichproben und gepaarter oder unabhängiger Messwerte unterschieden. Gepaarte Messwerte treten beispielsweise bei zeitlich verschobenen Untersuchungen an der gleichen Stichprobe, zum Beispiel Blutdruckwerte beim gleichen Patienten, oder Untersuchungen unter verschiedenen Umweltbedingungen, zum Beispiel Kohlenstoffaufnahme der gleichen Pflanze bei viel oder wenig  $\text{CO}_2$ , auf. Speziell im Hinblick auf den Beispieldatensatz können auch Simulationen mit Nudging als gepaarte Stichproben angesehen werden. So haben auch Righi et al. ?? in ihrer ursprünglichen Arbeit zur Analyse des Datensatzes mit der Annahme gepaarter Stichproben gearbeitet, auch wenn dies nicht explizit erwähnt wurde.

Die allgemeine Form der Teststatistik  $t$  von  $t = \frac{Z}{s}$  mit  $Z = Z(\bar{X}, \dots)$  und  $s = s(\text{Var}(X), \dots)$  als Funktionen der Daten wird durch die Form von  $Z$  und  $s$  im Detail je nach Situation modifiziert. Zum Beispiel gilt bei einem gepaarten t-Test mit dem Mittelwert  $\bar{X}_D$  und der Standardabweichung  $s_D$  der paarweisen Differenzen  $t = \frac{\bar{X}_D - \mu_0}{s_D / \sqrt{n}}$ . Auch die zur Berechnung des p-Wertes mit der t-Verteilung zusätzlich benötigte Anzahl  $\nu$  der Freiheitsgrade differiert je nach Situation; im Fall der gepaarten Differenzen ist  $\nu = n - 1$  bei  $n$  beobachteten Paaren. Die Nullhypothese besteht in der Annahme von keinem Unterschied zwischen den Stichproben beziehungsweise der Gleichheit mit einem festen Wert im Ein-Stichproben-Fall. Wird die Nullhypothese also signifikant abgelehnt, so wird den Daten ein Unterschied voneinander beziehungsweise von dem in  $H_0$  festgelegten Wert attestiert.

Voraussetzungen für die Anwendung eines t-Tests sind die Unabhängigkeit von  $Z$  und  $s$ , die Normal-Verteilung von  $X$  und die  $\chi^2$ -Verteilung von  $s^2$ . Alternativen für nicht-Normalverteilte Daten sind der Mann-Whitney U-Test (bei gepaarten Werten: Wilcoxon-Vorzeichen-Rang-Test) und bei ungleichen Varianzen in den zwei Stichproben der Welch'sche t-Test. Die Wahl der korrekten Tests fällt besonders bei kleinen Stichproben ins Gewicht.

### 3.1.4. Tests auf Normalverteilung

Voraussetzung für die Anwendung eines t- bzw. Welch-Tests für den Mittelwertvergleich ist die Normalverteilung der Daten. Tests auf Normalverteilung gibt es viele, beruhend auf Simulationsstudien wird häufig der Shapiro-Wilk-Test empfohlen. Dieser besitzt unter verschiedenen Situationen mit die größte Power [75]. Ähnlich gut schneidet aber auch der Anderson-Darling-Test oder der D'Agostino-Pearson Test ab. Dabei muss bemerkt werden, dass die Teststärke mit der Stichprobengröße im Allgemeinen steigt und daher vor allem für kleine Stichproben ( $n < 20$ ) die Normalverteilungsannahme selten abgelehnt wird [2]. Im Gegensatz zu vielen anderen Testprozeduren ist beim Normalverteilungstest die Nullhypothese die Annahme einer Normalverteilung. Ist das Ergebnis also die Feststellung einer signifikanten

Ablehnung der Nullhypothese, so wird damit den Daten eine Abweichung von der Normalverteilung zugeschrieben.

## 3.2. Hypothesentests mit dem Beispieldatensatz

Bei der Fragestellung, ob sich die beiden Simulationen signifikant unterscheiden, wird implizit davon ausgegangen, dass der Unterschied insgesamt festgestellt werden soll, also durch eine einzelne Entscheidung beziehungsweise Maßzahl. Dies kann durch Testen der globalen Mittelwerte geschehen, dann geht aber die räumliche Information vollständig verloren. Alternativ kann die multivariate Variante des t-Tests, der Hotelling- $T^2$ -Test, verwendet werden. Dies ist allerdings mit sehr großer Rechenzeit verbunden, wenn der zu testende multivariate Vektor  $\mathbf{x} = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_{155648})$  den zeitlichen Verlauf oder zeitliche Mittelwerte pro Pixel für alle 155.648 räumlichen Punkte des Datensatzes enthält. Alleine bei Anwendung auf eine einzelne Höenschicht muss die Kovarianzmatrix aller 8192 Pixel berechnet und invertiert werden. Verwendet man statt dessen die einzelnen Zeitpunkte als den zu testenden Vektor, also  $\mathbf{x} = (x_1, x_2, \dots, x_{120})$  mit der räumlichen Verteilung  $\vec{x}_i$ , ist die Signifikanz durch die Stichprobengröße von alleine  $n = 8192$  bei Betrachtung einer einzelnen Höenschicht quasi garantiert. Möchte man statt dessen pixelweise testen, so kann danach durch eine Multiples-Testen-Korrektur die Einhaltung des globalen Signifikanzniveaus sichergestellt werden.

Sämtliche Mittelwerts-Testprozeduren gehen von untereinander unabhängigen Daten aus. Da bei allen hier betrachteten Variablen sowohl eine räumliche als auch eine zeitliche Korrelation festgestellt wurde, müssen diese erst entfernt werden, bevor Tests ausgeführt werden können oder sie müssen in der Testprozedur bereits mit berücksichtigt werden. Die zeitliche Korrelation kann durch einen entsprechend groß gewählten Abstand zwischen den verwendeten Zeitpunkten verringert werden. Es werden zwei Möglichkeiten miteinander verglichen: Entweder eine Stichprobe besteht pixelweise aus den jeweiligen Monatswerten (also zum Beispiel alle Januar-Werte von 1996 bis 2005) oder aus den Jahresmittelwerten des gesamten Zeitraumes. Bei beiden Möglichkeiten ist die Größe der (pixelweisen) Stichprobe  $n = 10$  mit insgesamt 8192 Pixeln.

Die grafischen Darstellungen der Testergebnisse sind in C.2 zu finden, ein Vergleich der Anzahl an signifikanten Pixeln ist in Tabelle C.2.4 aufgelistet. Die folgenden Aussagen beziehen sich, soweit nicht explizit anders angegeben, auf die Differenzen aller betrachteter Variablen in der untersten Höenschicht.

### 3.2.1. Test auf Normalverteilung und Auswahl der t-Testprozedur

Als erstes wird die Normalverteilungsannahme für die Daten überprüft. Da aber die Stichprobe nicht besonders groß ist ( $n = 10$ ) und das Ergebnis des Shapiro-Wilk-Tests damit mit Vorsicht zu interpretieren ist, wäre es sicherer, gleich anstelle eines parametrischen Tests (hier: Welch-t-Test) die nicht-parametrische Alternative (hier: Wilcoxon-Vorzeichen-Rang-Test) zu verwenden. Es stellte sich im Nachhinein heraus, dass in diesem Fall beide Tests (parametrisch und nicht-parametrisch) zu fast identischen Ergebnissen geführt haben. Unterschiede haben sich nur im

Verhalten in Bezug auf die FWER-Korrekturen ergeben. Dies ist konsistent mit dem Ergebnissen der Shapiro-Wilk-Tests, bei welchen sich herausgestellt hat, dass die Normalverteilungsannahme bei den meisten Pixeln nicht abgelehnt werden kann. Die meisten Ausnahmen waren in Nähe des Äquators und über Landflächen zu finden. Auch der Absolutbetrag der Differenzen folgte diesem Muster, außer dass sich bei  $\text{NO}_3$  und dem Wolkenbedeckungsgrad eine Abhängigkeit vom Jahresverlauf zeigte. Aufgrund dieser leichten verbleibenden Unstimmigkeiten ist der nicht-parametrische Test vorzuziehen.

### 3.2.2. Vergleich der t-Tests

Es stellt sich die Frage, ob statt einem zweiseitigen Test der Differenzen ein einseitiger Test des Absolutbetrags der Differenzen anwendbar ist, da der einseitige Test stets eine größere Teststärke aufweist als der zweiseitige. Ein zweiter Vorteil eines einseitigen Tests besteht darin, dass damit nicht nur die strikte Abweichung von Null untersucht werden kann, sondern durch die Festlegung einer zwar kleinen, aber positiven Schranke  $\epsilon$  auch kleine Abweichungen von der Null ausgeschlossen werden können. Dies ist interessant, da von zweiseitigen Tests auch eine geringe Differenz, zum Beispiel in der fünften Nachkommastelle, als signifikant erkannt werden kann, obwohl diese physikalisch gesehen nicht von Bedeutung ist. Es wurden zwei Möglichkeiten, eine solche kleine Schranke festzulegen, untersucht und mit der einseitigen und zweiseitigen Formulierung verglichen. Insgesamt wurden also vier verschiedene Hypothesen überprüft:

1.  $H_0 : \mu = 0$ : zweiseitiger Test
2.  $H_0 : \mu \leq 0$ : einseitiger Test
3.  $H_0 : \mu \leq \epsilon_q$ : einseitiger Test mit Schranke  $\epsilon_q = \text{globales } 25\text{-Quantil}$
4.  $H_0 : \mu \leq \epsilon_d$ : einseitiger Test mit Schranke  $\epsilon_d = \text{globales Mittel}/100$  (entspricht ungefähr der zweiten Nachkommastelle)

Betrachtet wird das Verhalten der Testergebnisse ohne Korrektur für multiples Testen. Dabei werden bei allen vier Variablen vom strengen einseitigen Tests mehr Nullhypothesen abgelehnt als beim zweiseitigen Test. Die Pixel mit den von den beschränkten einseitigen Tests nicht abgelehnten Nullhypothesen liegen größtenteils an den Polen, wohingegen die vom zweiseitigen Test abgelehnten Pixel häufig über den großen Landflächen der Kontinente liegen. Der einseitige Test mit Null als Grenze führt zu abgelehnten Nullhypothesen bei allen Pixeln.

### 3.2.3. Vergleich der Multiplen-Testen-Korrekturen

Es wurde untersucht, wie sich die unkorrigierten p-Werte im Vergleich zur Anwendung der FWER-Korrekturen (Bonferroni, Holm, Hochberg und Hommel), der Benjamini-Hochberg-Korrektur und der Benjamini-Yekutieli-Korrektur (beide FDR-Kontrolle) verhalten. Dabei zeigt sich innerhalb der Gruppe der FWER- und FDR-kontrollierenden Methoden eine große Ähnlichkeit der räumlichen Verteilung der Pixel mit als signifikant abgelehnten Nullhypothesen. Zwischen den beiden

Gruppen und den unkorrigierten Ergebnissen sind jedoch teils erhebliche Unterschiede erkennbar. Die wenigsten Pixel mit abgelehnten Nullhypothesen erhält man bei Anwendung der FWER-kontrollierenden Korrekturen, da diese das globale  $\alpha$  sehr konservativ abschätzen. Die Benjamini-Hochberg-Korrektur führt zu fast identischen Ergebnissen wie die unkorrigierten Tests, außer beim Wolkenbedeckungsgrad. Die räumliche Verteilung der Ergebnisse der Benjamini-Yekutieli-Korrektur sieht meist wie ein Mittelwert zwischen den Ergebnissen von FWER-Korrektur und Benjamini-Hochberg-Korrektur. Die Abbildung 3.1 ist dabei repräsentativ für die meisten der durchgeführten Tests.

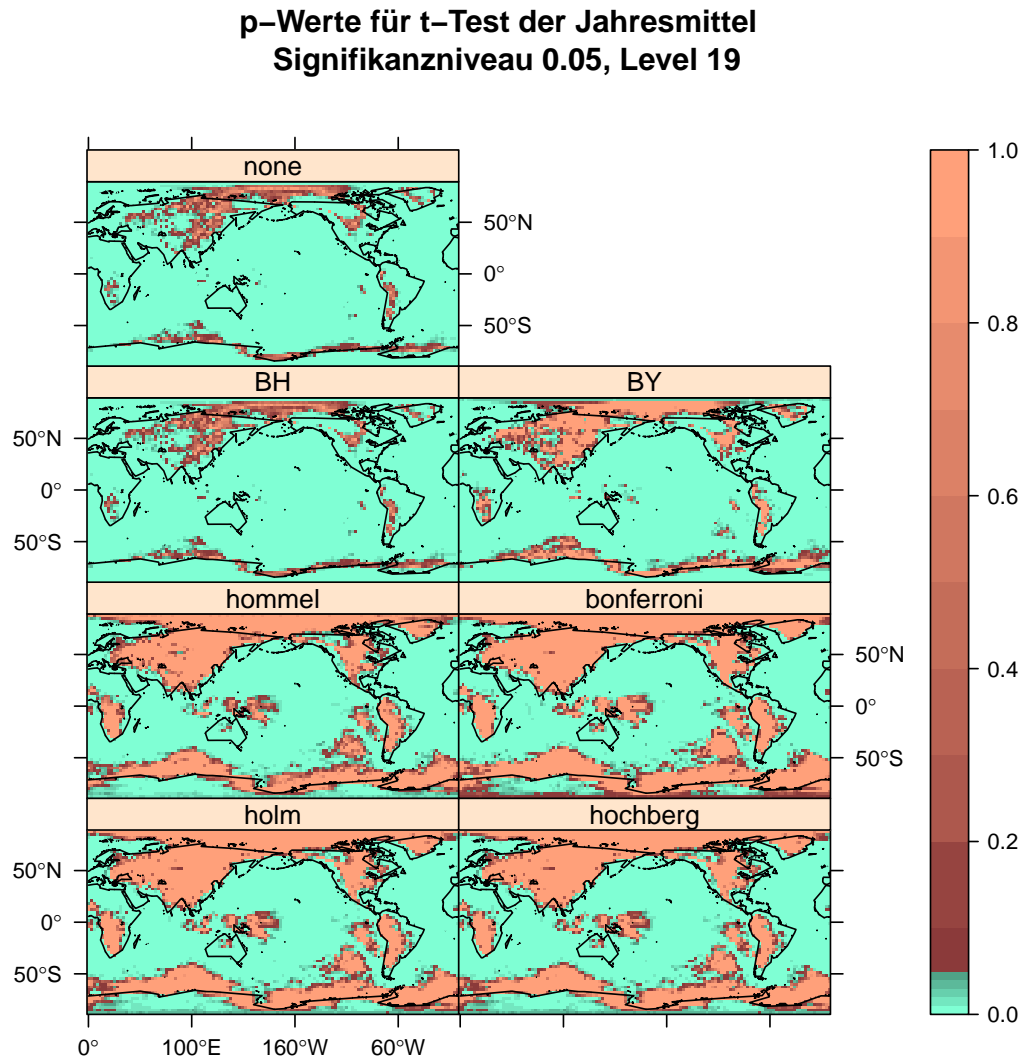


Abb. 3.1.: Räumliche Verteilung der p-Werte für einen zweiseitigen t-Test auf den Jahresmittelwerten der Differenz der Sulfat-Konzentration in der untersten Höschicht. Wird ein Signifikanzniveau von 5% angenommen, so zeigt die türkise Farbe eine Ablehnung der Nullhypothese zu diesem Niveau an, was hier eine Ablehnung der Normalverteilungsannahme bedeutet.

### 3.2.4. Ergebnisse der korrigierten Hypothesentests

Zur Auswertung der Jahresmittel-Differenzen in der Konzentration von  $\text{SO}_4$ ,  $\text{NO}_3$ , Ruß (BC) und dem Wolkenbedeckungsgrad wurde ein globales Signifikanzniveau von  $\alpha = 0.05$  festgelegt und der zweiseitige Wilcoxon-Vorzeichen-Rang-Test verwendet. Als bevorzugte Korrekturmethode für das Multiple Testen wird die Benjamini-Yekutieli-Korrektur betrachtet, da diese sowohl unabhängig von der Assoziationsstruktur der Hypothesen ist, als auch nicht so streng wie die FWER-kontrollierenden Methoden.

Die Konzentrationsdifferenzen von Ruß werden bei 80% aller Pixel als signifikant von Null verschieden erkannt. Die nicht signifikanten Pixel liegen über fast ganz Asien, Südamerika, Teilen Nordamerikas und der südlichen Hälfte von Afrika. Nur wenige Pixel liegen nicht über Landflächen, sondern vereinzelt im südlichen Atlantik, im Indischen Ozean, vor der chilenischen Küste und in der Antarktis. Auch bei  $\text{NO}_3$  (70% signifikant) liegen die meisten der nicht signifikanten Pixel über den Landflächen von Afrika, Südamerika, Sibirien, Indien und Australien. Zusätzlich wird fast die gesamte Arktis, der mittlere Atlantik und der östliche tropische Pazifik von der Signifikanz ausgeschlossen. Ebenfalls stark auf Asien konzentrieren sich die nicht signifikanten Pixel der  $\text{SO}_4$ -Differenz (81% signifikant). Bis auf wenige weitere Landflächen (Chile, Nordamerika, Südafrika, Grönland), wo vereinzelt nicht signifikante Pixel vorhanden sind, verteilt sich der Rest auf die Arktis und die Küsten vor der Antarktis.

Bei allen drei Konzentrationsdifferenzen zeigte sich bei Verwendung der einzelnen Monate als Stichprobe ein Jahresgang in der räumlichen Verteilung der signifikanten Pixel.

Davon komplett verschieden ist der Wolkenbedeckungsgrad, wo nicht bei einem einzigen Pixel die Nullhypothese abgelehnt wird. Hätte man in diesem Fall keine Korrektur für Multiples Testen angewendet, so wären circa 4.5% der Pixel als signifikant gekennzeichnet worden, was ungefähr der Fehlerrate  $\alpha$  entspricht. Hier zeigt sich die größere Trennschärfe bei Verwendung einer Korrektur. Repräsentativ ist in Abbildung A.12 das Ergebnis des Wilcoxon-Vorzeichen-Rang-Tests für die Differenz des Wolkenbedeckungsgrades der untersten Höschicht zu sehen.

## 3.3. Anwendungsbereich von Hypothesentests

Das Verfahren des statistischen Hypothesentests besitzt ein stabiles und lange erprobtes theoretisches Fundament. Dies führt zu klaren Anweisungen zur Testdurchführung und leicht überprüfbaren Voraussetzungen zur Anwendung der Testprozedur. Zur Entscheidung für eine Testprozedur müssen Verteilungsvoraussetzungen an den Daten geprüft werden und eine zur Problemstellung passende Hypothese (einseitig oder zweiseitig) formuliert werden. Wenn eine Aussage über ein globales Signifikanzniveau getätigt werden soll (zum Beispiel beim Vergleich zweier Felder/Bilder, auch 3D), ist es empfehlenswert, zusätzlich über eine Korrektur für Multiples Testen nachzudenken. Die Durchführung des Tests und eventueller p-Wert-Korrekturen an sich ist nur mit geringem Zeitaufwand verbunden.

Als zwei offensichtliche Nachteile der Verwendung von Hypothesentests hat sich an dem hier betrachteten Beispiel die geringe Sensitivität für kleine Unterschiede und die Voraussetzung unabhängiger Datenpunkte in der Stichprobe heraus-

gestellt. Ebenfalls hier, aber auch universell ungünstig, ist die Abhängigkeit der Signifikanz von der Größe der Stichprobe. Als Alternative oder Ergänzung wird oft die (zusätzliche) Angabe der Effektgröße, beispielsweise gemessen durch ein Effektmaß wie Cohen's  $d$ , empfohlen.

Eine andere Folge des ausführlichen theoretischen Fundaments ist allerdings, dass die Frage nach der Verwendung von Multiplen-Test-Korrekturen oder überhaupt von Hypothesentests in mathematisch-philosophisch-wissenschaftstheoretische Grundsatzdiskussionen führt, welche bereits in verschiedenen wissenschaftlichen Disziplinen geführt werden, aber noch zu keiner abschließenden Handlungsvorschrift geführt haben. Eine kurze Übersicht des aktuellen Stands der Diskussion ist in den Kommentaren von Thompson [66] zu [60] zusammengefasst. Die grundlegende Empfehlung zur Umgehung dieser inhärenten Problematik besteht in der Verwendung von (Semi-)Bayesianischen Methoden, wovon ein paar in ähnlicher Anwendung in Kapitel 5 und 7 angesprochen werden.



# 4. Orthogonalzerlegung PCA (EOF)

## 4.1. Theoretischer Hintergrund

### 4.1.1. Beschreibung des Verfahrens

Die Hauptkomponentenanalyse ist in der Statistik unter der Abkürzung PCA (Principal Components Analysis) bekannt und so auch in vielen Büchern über Multivariate Analysemethoden, zum Beispiel Konishi (2014) [35] oder Backhaus et al. (2000) [6] zu finden. In der Geostatistik und der Meteorologie wird das Verfahren auch als EOF-Analyse (EOF = Empirical Orthogonal Functions) bezeichnet. Eine ausführliche Erläuterung speziell im Bezug auf die Anwendung in der Atmosphärenphysik findet sich in von Storch und Zwiers (2002) [69], Anwendungsmöglichkeiten und Beispiele werden in von Storch und Navarra (1995) [68] diskutiert. Die folgende kurze Erklärung ist übernommen aus [35, 69].

Mit einer PCA können zwei Ziele verfolgt werden. Die Daten werden über eine Singulärwertzerlegung auf einen neuen Basisraum projiziert. Dabei sind die neuen Basisvektoren orthogonal zueinander und die Projektionen der Daten auf die neuen Basisvektoren, die Hauptkomponenten, beschreiben jeweils einen maximalen Anteil an Varianz in den Originaldaten. Dadurch können Strukturen in den Daten gefunden werden, die vor der Projektion nicht offensichtlich waren, weshalb die PCA auch zu den strukturentdeckenden Verfahren der multivariaten Analyse gezählt wird. Zusätzlich können die Dimensionen der Daten ohne größeren Informationsverlust reduziert werden, falls durch wenige Hauptkomponenten ein Großteil der Varianz abgedeckt ist.

Betrachtet man Originaldaten der Form  $\mathbf{x} = (x_1, x_2, \dots, x_p)$ , mit jeweils  $n$  Beobachtungen in  $x_i$ , so bewirken die Bedingung der Orthogonalität und die Maximierung des Varianzanteils, dass die neuen Basisvektoren des Projektionsraumes die Eigenvektoren  $w_i$  der Kovarianzmatrix  $\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$  der Originaldaten sind. Die neuen Basisvektoren  $w_i$  werden also durch eine Eigenwertzerlegung  $\mathbf{S} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T$  der Kovarianzmatrix berechnet. Dabei entspricht die Maximierung des Varianzanteils dem Sortieren nach der Größe des zum jeweiligen Eigenvektor gehörigen Eigenwertes  $\lambda_i$ , aus welchem sich auch der Anteil  $p$  der erklärten Varianz  $p_i = \frac{\lambda_i}{\sum_{j=1}^N \lambda_j}$  berechnet. Die zu jedem Eigenvektor gehörende Hauptkomponente ergibt sich aus der Projektion  $a_i = w_i^T \mathbf{x}$ . Die bei der Anwendung auf ein physikalisches Problem auftretenden Einheiten der ursprünglichen Variablen werden, soweit vorhanden, bei der Zerlegung auf die Hauptkomponenten übertragen [69]. Diese können dann mit einem multivariaten Test, wie zum Beispiel Hotellings- $T^2$ -Test für Differenzen, auf Verschiedenheit von Null getestet werden [68].

In der Statistik entsprechen die einzelnen Variablen  $\mathbf{x}_i$  häufig verschiedenen Eigenschaften eines Objektes, wie zum Beispiel Größe oder Farbe. In der Geostatistik und Meteorologie werden die zeitlichen oder räumlichen Koordinaten als 'Eigenschaft' verwendet. Meist ist es von Interesse, die zeitliche Entwicklung (= Hauptkomponenten) von räumlichen Mustern (= Eigenvektoren, EOFs) zu betrachten, wobei die umgekehrte Formulierung mathematisch gesehen genauso möglich wäre. Weil die PCA sensitiv auf absolute Größenunterschiede ist, wird bei der gemein-

samen Analyse von Variablen mit verschiedenen Einheiten vor Berechnung der Eigenvektoren häufig eine Zentrierung (auf Mittelwert 0) oder Standardisierung vorgenommen. Da im Fall räumlicher oder zeitlicher Aufteilung nur eine Variable betrachtet wird, ist die Verwendung der Zentrierung beziehungsweise Standardisierung nicht zwingend vorgegeben.

Die PCA ist nur für normalverteilte Daten in dem Sinne ein optimales Verfahren, dass die resultierenden Hauptkomponenten vollständig unkorreliert sind [69]. Für den Fall nicht-normalverteilter Daten gibt es das Verfahren der Independent Component Analysis (ICA) aus dem Bereich der Signalverarbeitung, welches eine optimale Trennung von additiv vermischten, nicht-normalverteilten und voneinander (statistisch) unabhängigen Quellen zum Ziel hat [31]. Der Begriff Quellen wird in der ICA ähnlich dem der Eigenvektoren in der PCA verwendet. Eine PCA kann keine dynamischen, in der zeitlichen Dimension propagierende, räumliche Muster erkennen [69]. Falls dies gewünscht ist, bietet sich die Verwendung der Hilbert-EOF-Analyse (auch komplexe EOF-Analyse), der erweiterten EOF-Analyse (EEOF) oder der Principal-Oscillation-Pattern-Analyse (POP) an [69, 16, 68]. Eine ausführliche Beschreibung und Diskussion findet sich in Cressie und Wikle (2011) [16].

#### **4.1.2. Implementierungsmöglichkeiten des Verfahrens**

Für die Berechnung von Eigenwertproblemen gibt es viele verschiedene numerische Verfahren. Daher werden im folgenden Abschnitt kurz die für die Analyse des Beispieldatensatzes verwendeten Algorithmen genauer vorgestellt. Die PCA wurde zweimal durchgeführt, einmal unter Verwendung des in der Meteorologie breit genutzten Software-Pakets *cdo* (Climate Data Operators) des Max-Planck-Instituts für Meteorologie [14] und ein zweites Mal mit dem Statistik-Software-Paket R [50] unter Verwendung des Zusatzpaketes *spacetime* [46]. Für die ICA wurde die in R implementierte Version [42] des *fastICA*-Algorithmus von Hyvärinen und Oja [31] verwendet.

##### **4.1.2.1. PCA**

Die Software-Sammlung *cdo* verwendet laut Dokumentation zur Berechnung der EOF-Analyse einen einseitigen Jacobi-Algorithmus, wie zum Beispiel im Handbuch für lineare Algebra [1] beschrieben. Dabei wird die Iteration bei der numerischen Eigenwertzerlegung direkt mit der Inputmatrix durchgeführt. Im Gegensatz dazu wird für den von R verwendeten Divide-and-Conquer-Algorithmus die Inputmatrix vor Beginn der Iteration erst bidiagonalisiert. Der Jacobi-Algorithmus ist am stabilsten, wenn die Inputmatrix diagonaldominant ist, das heißt die Werte in der Diagonalen oder nahe daneben wesentlich größer sind als die abseits der Diagonalen.

##### **4.1.2.2. ICA**

Der *fastICA*-Algorithmus besteht aus drei Schritten. Im erste Schritt werden die Daten zentriert, also der jeweilige Mittelwert abgezogen. Im zweiten Schritt werden die Daten durch Anwendung einer PCA dekorreliert. Im dritten Schritt werden

iterativ die einzelnen Quellen (entsprechend den Eigenvektoren bei der PCA) extrahiert. Die zu maximierende Größe ist ein Maß für die nicht-Normalverteilung der Projektionen, zum Beispiel die Kurtosis (= Verteilungskenngröße zum Messen des Gewichts der Extrema oder die Spitzheit der Verteilung). Der in *fastICA* implementierte Algorithmus verwendet die Negentropie  $J$  aus der Approximation  $J(y) = (E[G(y)] - E[G(v \sim N(0, 1))])^2$  als Maß für die Ähnlichkeit der Daten  $y$  zu einer Standardnormalverteilung  $v$ ,  $G$  ist dabei eine zu Null symmetrische u-förmige Funktion.

## 4.2. PCA mit dem Beispieldatensatz

Es wurde am Beispiel der Sulfat-Konzentrationsdifferenz der bodennächsten Schicht untersucht, ob die PCA in der Lage ist, aus den Jahresmittelwerten, den jeweiligen gleichartigen Monaten (alle Januare, alle Februar, ...) und dem kompletten Datensatz (sämtliche 120 Monate) räumliche Muster zu finden, deren Hauptkomponenten sich signifikant von Null unterscheiden. Es werden die Hauptkomponenten und nicht die Eigenvektoren getestet, da die physikalischen Einheiten auf die Hauptkomponenten übertragen werden. Als Test wurde die multivariate Variante des t-Tests, der Hotelling- $T^2$ -Test, verwendet. Dabei wurde zusätzlich untersucht, ob die Signifikanz des Tests von der Anzahl der verwendeten Hauptkomponenten abhängt. Da dies zum Vergleich mehrerer Hypothesen an der gleichen Stichprobe führt, wurde auch hier eine Korrektur für das multiple Testen angewendet. Sämtliche grafischen Darstellungen und die Tabellen mit den p-Werten des Hotelling- $T^2$ -Tests sind in ?? zu finden.

### 4.2.1. Vergleich der Verwendung von Jahresmittelwerte, gleichartigen Monaten und dem kompletten Zeitverlauf

Im Folgenden wird die PCA der Differenz der Sulfat-Konzentration ausführlicher beschrieben. Die Analyse der übrigen Variablen folgt dem gleichen Schema.

#### 4.2.1.1. EOF der Jahresmittelwerte und gleichartigen Monate der Konzentrationsdifferenz

**Unveränderte Jahresmittelwerte** Unabhängig von der betrachteten Höhengschicht erklärt der erste Eigenvektor den Großteil der Varianz. Auch die räumliche Verteilung, wie beispielhaft in Abbildung 4.1 zu sehen ist, unterscheidet sich von der der restlichen Eigenvektoren. Es treten wesentlich großflächigere Strukturen mit durchwegs gleichem Vorzeichen auf. In den weiteren Eigenvektoren sind die Strukturen mit gleichem Vorzeichen viel kleinskaliger. Die einzigen zwei Vorzeichenwechsel beim ersten Eigenvektor in den unteren Höhengschichten finden sich in den beiden Polregionen, wo die Differenz der Sulfat-Konzentration global gesehen am geringsten ist. Je höher die betrachtete Schicht ist, desto weiter wandert die Grenze des Vorzeichenwechsels von der Antarktis nach Norden und in den oberen Schichten tritt kein Vorzeichenwechsel mehr auf. Das räumliche Muster in den unteren Schichten zeigt (zumindest grafisch) Ähnlichkeiten zum Schiffsemissionsmuster. Die zugehörige Hauptkomponente des ersten Eigenvektors ist in allen

Schichten (bis auf die obersten drei) eindeutig positiv, was sich auch in den Ergebnissen des Hotelling- $T^2$ -Tests widerspiegelt. Dieser erste Eigenvektor sorgt dafür, dass der Hotelling- $T^2$ -Test mit bis zu sechs zusätzlichen, also insgesamt den ersten sieben Eigenvektoren signifikant zum Niveau  $\alpha = 0.05$  ist. Werden die restlichen Eigenvektoren ohne den ersten, also Nummer 2 bis 10, getestet, so stellt sich für keine der möglichen Kombinationen eine Signifikanz ein.

**Zentrierte und skalierte Jahresmittelwerte** Hier erklärt der erste Eigenvektor nur noch in den oberen vier Schichten mehr als 50% der Varianz, in den unteren Schichten benötigt man sogar die ersten neun Eigenvektoren, um 50% zu erklären. Dies spiegelt sich auch in den Ergebnissen des Hotelling-Tests wieder, wo alle Kombinationen bis auf solche mit sehr vielen Eigenvektoren zu 0.05 signifikant sind. Die räumliche Struktur der Eigenvektoren unterscheidet sich nicht stark voneinander; es sind in den unteren Schichten weder großskalige noch sehr kleinskalige Bereiche ohne Vorzeichenwechsel zu sehen. Je höher die Schichten liegen, desto großskaliger und räumlich homogener werden die positiven und negativen Bereiche. Keines der Muster ist (zumindest visuell) mit dem der Schiffsemissionen in Verbindung zu setzen.

**Gleichartige Monate** Im Vergleich zu den Ergebnissen der Jahresmittel-EOF-Analyse zeigen die Ergebnisse der Analyse der jeweils gleichartigen Monate ein etwas variableres Verhalten. Die Eigenvektoren weisen eine etwas höhere räumliche Variabilität auf und der Anteil an erklärter Varianz ist für die ersten Eigenvektoren etwas geringer. Die größere Unsicherheit zeigt sich auch in den Ergebnissen des Hotelling-Tests. Während bei der Analyse der unveränderten Konzentrationsdifferenz nach wie vor die erste Hauptkomponente zu  $\alpha = 0.05$  signifikant von Null verschieden ist, so ist dies bei den Analysen der zentrierten und standardisierten Differenzen nur für den August der Fall.

#### 4.2.1.2. EOF des gesamten Zeitverlaufs der Konzentrationsdifferenz

**Unveränderte Differenzen** Hier erklärt der erste Eigenvektor vor allem in den untersten und obersten Höenschichten durchschnittlich 75% der Varianz, in den mittleren Schichten werden zur Erklärung des gleichen Anteils bis zu 40 zusätzliche Eigenvektoren benötigt. Mit den ersten beiden Eigenvektoren zusammen lassen sich 80% der Varianz erklären, möchten man einen Anteil von über 90% erreichen, so müssen ungefähr die ersten 20 Eigenvektoren betrachtet werden. Die räumliche Verteilung der Eigenvektoren zeigt wieder eine Homogenisierung für die höheren Schichten. Auch sind in den ersten fünf bis sechs Eigenvektoren wesentlich großskaligere Flächen von einheitlichen Vorzeichen vorhanden. Bei den untersten vier bis fünf Schichten ist im ersten Eigenvektor äußerst deutlich das Muster der Schiffsemissionen zu erkennen (exemplarisch in Abbildung A.13), aber auch in der zweiten Komponente zeigen sich Ähnlichkeiten zu den Schiffsmustern. Allerdings können auch in den weiteren Eigenvektoren Teilestrukturen des Emissionsmusters wiedererkannt werden. Der Hotelling-Test der Hauptkomponenten ergibt eine Signifikanz der ersten und achten Hauptkomponente, aber auch Kombinationen aus mehreren Hauptkomponenten erreichen bei einem  $\alpha$  von 0.05 ein signifikantes Testergebnis. Der zeitliche Verlauf der ersten Hauptkomponenten, zu sehen für die unterste Hö-

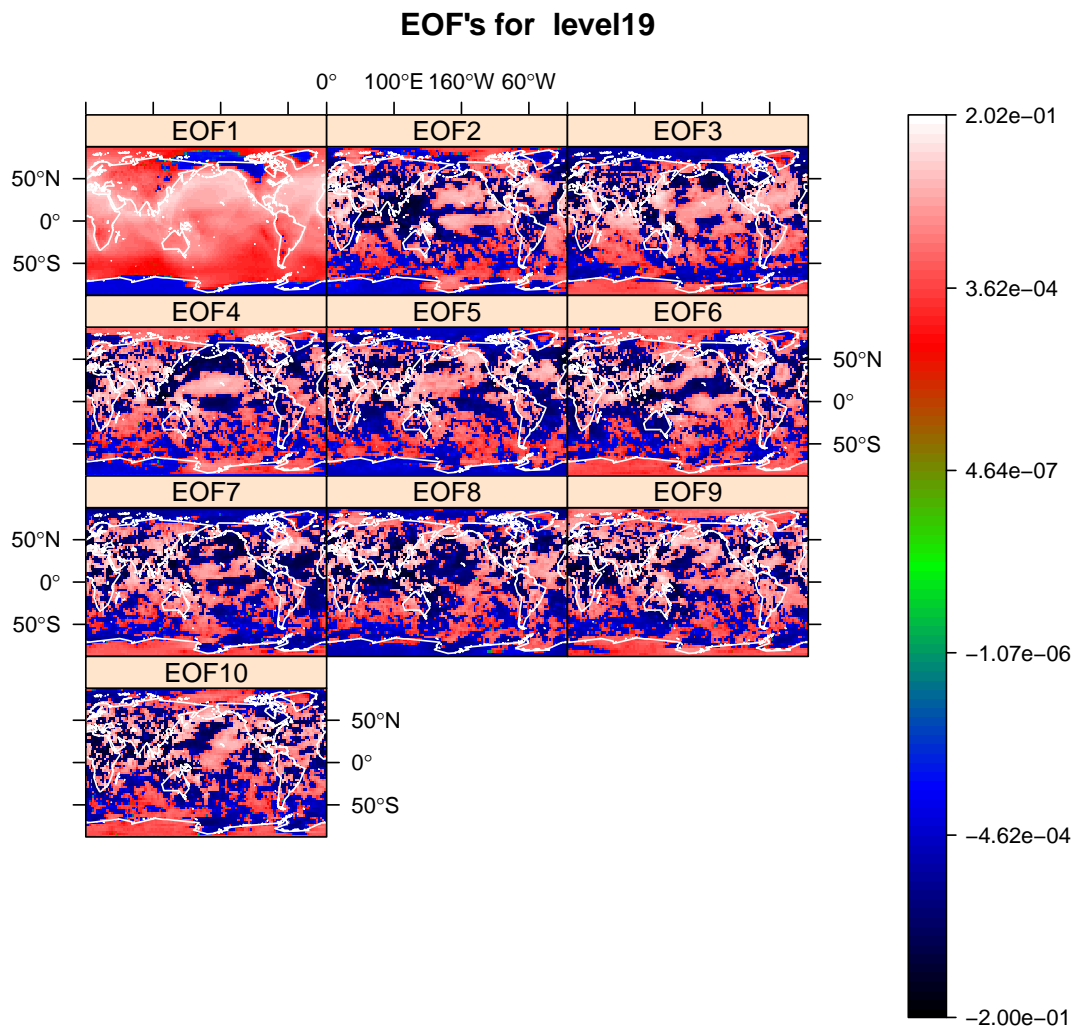


Abb. 4.1.: Räumliche Eigenvektoren der Jahresmittelwerte der Sulfat-Differenz der untersten Höschicht

henschicht in Abbildung A.14, zeigt im Gegensatz zu dem der Jahresmittel- und Monatsanalysen eine größtenteils regelmäßige Schwankung im Jahresverlauf, besonders die ersten beiden Hauptkomponenten der unteren Schichten. Dabei zeigen sie manchmal eine gegenläufige Schwankung, wobei die erste Hauptkomponente stets im Sommer ihr Maximum hat.

**Zentrierte und standardisierte Differenzen** Im Unterschied zur Analyse der Jahresmittel- und gleichartigen Monatswerte kann hier auch bei Zentrierung und Standardisierung in den ersten Eigenvektoren (1 bis ca. 6) eine Ähnlichkeit zu den Schiffsemissionsmustern festgestellt werden (Abbildung C.3.2). Das Ergebnis des Hotelling- $T^2$ -Tests ist mit zwei Ausnahmen für alle möglichen Kombinationen von Hauptkomponenten signifikant. Die Schwankung der Hauptkomponenten im Jahresrhythmus zeigt sich bei der Analyse der skalierten Differenzen durchaus bis zur fünften Hauptkomponente, wobei teilweise eine Überlagerung mit einer halbjährlichen Schwankung zu erkennen ist. Bei den Hauptkomponenten der zentrier-

ten Differenzen zeigt sich der Jahresgang nur in der ersten und die halbjährliche Schwankung in der zweiten Hauptkomponente. Die Ergebnisse des Hotelling-Tests zeigen durchwegs eine Signifikanz zu  $\alpha = 0.05$  an.

**Vergleich der Analysen** Die Tatsache, dass sich die ersten Eigenvektoren der Analyse der unveränderten der Jahresmittelwerte und des unveränderten gesamten Zeitverlaufes räumlich sehr ähnlich sind, führt zu der Interpretation, dass dieses Muster den globalen Durchschnitt beschreibt, vor allem da es bei der Analyse nach der Zentrierung der Differenzen nicht mehr zu finden ist. Die bei der Analyse des gesamten Zeitverlaufs nach der Zentrierung erkennbaren Muster beschreiben die zeitliche Variabilität, die natürlich bei Verwendung von Jahresmitteln oder gleichartigen Monaten vorher schon aus den Daten entfernt wurde. Die Extremwerte dieses ersten Eigenvektors liegen auf den Schifffahrtsrouten und weisen einen Vorzeichenwechsel zwischen Nordatlantik-/pazifik und Ozeanien/Indischer Ozean auf.

#### 4.2.2. EOF-Analyse für die anderen Variablen

Die Ergebnisse der Analysen der anderen beiden Konzentrationen zeigen in Bezug auf die Unterschiede bei Verwendung von unveränderten, zentrierten oder standardisierten Jahresmittelwerten, gleichartigen Monatswerten oder dem kompletten Zeitverlauf das gleiche Verhalten wie die Ergebnisse der Sulfat-Analyse. Die Ergebnisse der Analyse der unveränderten Ruß-Konzentrationsdifferenz zeigt nur im Verhalten der Hauptkomponenten einen Unterschied. Dort ist nur in der ersten Hauptkomponente ein Jahresgang erkennbar. Auch die Ergebnisse der Nitrat-Konzentrationsdifferenz-Analyse zeigen ein anderes Verhalten im zeitlichen Verlauf der Hauptkomponenten. Bei diesen sind für wesentlich mehr Hauptkomponenten regelmäßige Schwankungen erkennbar. Auch zeigt der erste Eigenvektor ein anderes räumliches Muster mit sehr wenig Vorzeichenwechsel, die ausschließlich über Land zu finden sind. Bei den räumlichen Mustern der Nitrat-Konzentrationen lassen sich nur sehr schwache Ähnlichkeiten mit dem Muster der Schiffsemissionen finden. Sämtliche Ergebnisse sind grafisch dargestellt in C.3.2.

Bei den Analysen des Wolkenbedeckungsgrades zeigte sich kein Unterschied in der Verwendung von unveränderten, zentrierten oder standardisierten Differenzen und auch kein Unterschied zwischen den Ergebnissen bei Verwendung der Jahresmittelwerte, der gleichartigen Monatswerte oder dem gesamten Zeitverlauf. Der Anteil der erklärten Varianz der Eigenvektoren zeigt fast eine Gleichverteilung auf den Eigenvektoren. Diese unterscheiden sich auch in ihrem räumlichen Muster nicht, welches einen sehr kleinskaligen Wechsel zwischen positiven und negativen Vorzeichen zeigt. Lediglich in den mittleren Höhengestirten zeigt sich über dem Äquator eine leichte Homogenisierung und ein leicht erhöhter Wert im Vergleich zu den Mittleren Breiten und den Polregionen. Für die untersten Schichten dreht sich dieses Verhalten um und die Mittleren Breiten weisen einen leicht erhöhten absoluten Wert auf. Die Hotelling- $T^2$ -Tests fallen allesamt nicht-signifikant aus.

### 4.3. ICA für den Beispieldatensatz

Bei der Analyse der PCA-Ergebnisse ohne Zentrierung fällt auf, dass die ersten beiden Hauptkomponenten ein genau entgegengesetztes Vorzeichen aufweisen und man könnte daher die These aufstellen, dass die Trennung in zwei Muster aufgrund der Orthogonalitätsbedingung in der PCA zustande kam. Die ICA lockert diese Bedingung zur statistischen Unabhängigkeit und maximiert statt dem Varianzanteil die nicht-Normalverteilung der Daten. Es wurde daher untersucht, ob die ICA in der Lage ist, die räumliche Ähnlichkeit der Differenz zu den Schiffsemissionen in einem einzelnen Muster darzustellen. Da für die ICA eine Vorauswahl per PCA getroffen wird, wurde ebenfalls untersucht, wie sich die Anzahl der verwendeten Eigenvektoren auf das Ergebnis auswirkt.

#### 4.3.1. ICA der Sulfat-Konzentrationsdifferenz der untersten Höhengschicht

Die Anwendung der ICA auf die unveränderten Jahresmittelwerte führte unabhängig von der Anzahl der verwendeten Eigenvektoren zu einem räumlich weniger homogen verteiltem Muster der Quellen (entspricht den Eigenvektoren der PCA). Bei der Anwendung auf die zentrierten und standardisierten Jahresmittelwerte war die ICA jedoch in der Lage bei Verwendung von bis zu fünf Eigenvektoren eine oder zwei räumlich homogener als die Eigenvektoren verteilten Quellen zu extrahieren, die eindeutig mit dem Schiffsverkehrsmuster in Verbindung zu bringen sind. Ein Vergleich für zwei verwendete Eigenvektoren findet sich in 4.2. Auch bei der ICA des vollständigen Zeitverlauf sind die extrahierten Quellmuster eindeutig dem Schiffsmuster ähnlich, zeigen aber keine Verbesserung gegenüber den Eigenvektoren der PCA in Hinblick auf ihre räumliche Homogenität oder Identifizierungsmöglichkeit mit dem Schiffsmuster (Grafiken in C.3.3). In den Projektionen, also sozusagen den Hauptkomponenten der ICA, ist keine regelmäßige Schwankung im Jahresverlauf erkennbar.

#### 4.3.2. ICA für die anderen Variablen in der untersten Höhengschicht

Die Grafiken, auf welchen die folgende Analyse beruht, finden sich in C.3.3. In den Quellmustern der Ruß-Analyse sind wesentlich deutlichere Verbindungen zum Schiffsemissionsmuster herzustellen. Bei der Anwendung auf die unveränderten Jahresmittel und den unveränderten kompletten Zeitverlauf hat sich im Gegensatz zur Sulfat-Analyse keine Verschlechterung der räumlichen Homogenität ergeben. Es findet auch eine klarere Herausarbeitung des Schiffsignals statt, das heißt dieses ist häufig nur in einer Quelle wiederzufinden und nicht teilweise noch in weiteren Quellmustern enthalten und zwar unabhängig von der Anzahl der eingangs verwendeten Eigenvektoren und ob Jahresmittel verwendet wurden oder nicht. Bei der ICA der Nitrat-Konzentrationsdifferenzen hingegen findet ähnlich wie bei der Sulfat-Analyse eine Aufteilung des räumlich homogenen ersten Eigenvektors auf mehrere Quellen statt und man kann keine Ähnlichkeit mit dem Emissionsmuster (unabhängig von der Anzahl der verwendeten Eigenvektoren und der zeitlichen

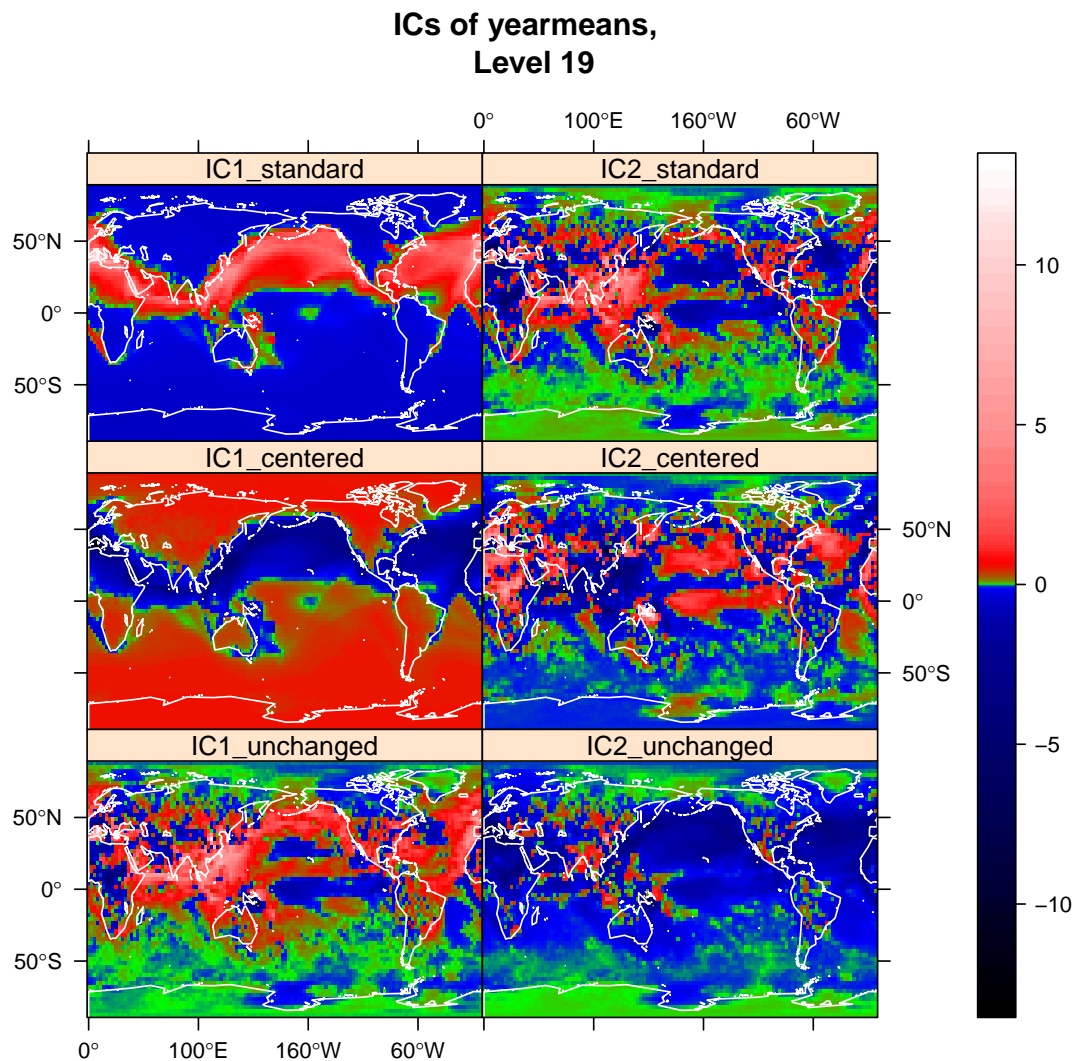


Abb. 4.2.: Räumliche Verteilung der Quellen der ICA der Sulfat-Differenz der untersten Hörschicht nach Selektion der ersten beiden Eigenvektoren der PCA

Mittelung) feststellen. Die ICA des Wolkenbedeckungsgrades hat zu keiner Veränderung im Vergleich zur PCA geführt.

#### 4.4. Vergleich PCA und ICA

In Hinblick auf die Klarheit der Unterscheidung zwischen dem Schiffsemissionsmuster und den restlichen Einflüssen war für die Analyse der Sulfat-Konzentrationsdifferenzen die PCA und für die Nitrat- und Ruß-Differenzen die ICA besser geeignet. Während die PCA nur bei Verwendung von unveränderten, also nicht zentrierten und standardisierten, Jahresmittelwerten und gleichartigen Monaten das Schiffssignal eindeutig erkannt hat, konnte unter Hinzunahme der ICA auch bei diesen Daten ein Emissionsmuster-ähnliches Muster gefunden werden. Bei der Analyse des Wolkenbedeckungsgrades konnten beide Methoden keine neuen Erkenntnisse erbringen, während bei den Analysen der Konzentrationsdifferenzen zumindest in



den Hauptkomponenten des unveränderten vollständigen Zeitverlaufs mindestens eine signifikante Hauptkomponenten gefunden werden konnte.

## 4.5. Anwendungsbereich PCA

Eine PCA wird als Transformation der Daten angewendet, um die in den Daten enthaltene Information zu kondensieren. Teile der Korrelationen innerhalb der Daten können beseitigt werden, die Größe des Datensatzes kann verringert werden und das Auffinden von die Varianz bestimmenden Strukturen erleichtert. Dadurch wurde bei dem hier besprochenen Beispielfall die Verwendung eines multivariaten t-Tests (Hotelling- $T^2$ ) möglich, welcher anders als die univariaten Varianten statt der iid-Vorraussetzung nur von einer multivariaten Normalverteilung ausgeht und somit Korrelationen zwischen den Datenvektoren, hier Pixeln, zulässt. Allerdings stellt sich bei der Dimensionsreduktion stets die Frage nach der Anzahl der beizubehaltenden Eigenvektoren. In der Statistik ist ein häufig verwendetes Kriterium, dass der Gesamtanteil der behaltenen Eigenvektoren an der erklärten Varianz einen bestimmten Wert überschreiten muss, welcher meist 80-90% beträgt. Auch die ICA beruht auf einer Auswahl an Eigenvektoren und steht damit vor dem gleichen Wahlproblem, hat sich aber größtenteils relativ robust gegenüber der Variation der Anzahl der verwendeten Eigenvektoren gezeigt. Falls also das interessierende Signal in Eigenvektoren verborgen ist, welche nicht wesentlich zur Varianzerklärung beitragen, so wird bei der Dimensionsreduktion durch die PCA (und damit auch der ICA) das Signal fälschlicherweise dem Rauschen zugeordnet und von der weiteren Analyse ausgeschlossen. Auch wenn das gesuchte Signal teilweise im Mittelwert enthalten ist, verschwindet es bei Zentrierung oder Standardisierung der Daten aus der weiteren Analyse. Sind also keine a priori Information über die räumliche Verteilung und zeitliche Variabilität des Signals vorhanden, kann in der EOF-Analyse leicht Information übersehen werden. Die physikalische Deutung der von der PCA gefundenen Eigenvektoren und Hauptkomponenten wird erschwert durch die Tatsache, dass Muster, welche einen großen Teil der Varianz in den Daten erklären, nicht zwangsweise mit den datenerzeugenden physikalischen Prozessen in Verbindung gebracht werden können [69]. Über mögliche Zusammenhänge zwischen Mustern und deren Stärke, zum Beispiel der Zusammenhang zwischen der ersten EOF der nicht zentrierten Sulfat-Konzentration und dem räumlichen Muster der Schiffsemissionen, ist durch eine PCA keine Aussage zu treffen.

# 5. Regressionsmodelle

## 5.1. Mathematischer Hintergrund

Regressionsmodelle werden verwendet, um einen Zusammenhang zwischen zwei Variablen zu quantifizieren. Die Art des Zusammenhangs kann dabei in Abhängigkeit der Eigenschaften der Variablen mit unterschiedlicher Komplexität modelliert werden. Die einfachsten Modelle sind lineare Modelle. Eine übersichtliche Einführung bietet Faraway (2005) [19], eine Spezialisierung auf die Anwendung in der Meteorologie findet sich in von Storch und Zwiers (2000) [69]. Auch in den meisten Büchern über komplexere Modelle ist meist eine Einführung in lineare Modelle enthalten. Die Beschreibung im folgenden Abschnitt ist übernommen aus Wood (2006) [71], aus dem auch die Erklärung der Generalisierten Additiven Modelle stammt.

Bei der Modellierung mit (univariaten) Regressionsmodellen wird der Einfluss einer oder mehrerer Größen (= Kovariablen  $\mathbf{X} = (x_1, x_2, \dots, x_n)$ ) auf eine Zielvariable  $y$  geschätzt. Dabei wird von einer additiven Struktur im Zusammenwirken der Kovariablen sowie von  $N(0, \sigma^2)$ -verteilten und unabhängigen Residuen  $\varepsilon$  ausgegangen, wobei daraus auch die Annahme einer Normalverteilung für die Daten  $y$  folgt. Dann kann das Modell folgendermaßen geschrieben werden:  $y = \hat{y} + \varepsilon = \beta\mathbf{X} + \varepsilon$  mit  $\beta\mathbf{X} = \beta_0 + \beta_1x_1 + \dots + \beta_nx_n$ . Im einfachsten Fall ist das Modell linear in den Koeffizienten  $\beta_i$  und auch in den Kovariablen. Zur Schätzung nicht-linearer Zusammenhänge können zum Beispiel nicht-parametrische Funktionen, dargestellt durch die Überlagerung von abschnittsweise an die Daten angepassten Basisfunktionen, verwendet werden (= Generalisierte Additive Modelle, GAM). Im Allgemeinen sind generalisierte Regressionsmodelle solche, wo die Normalverteilungsannahme für  $y$  gelockert wurde und dann auch andere Verteilungen, hauptsächlich aus der Exponentialfamilie, angenommen werden können. Für beliebige Verteilungsannahmen können GAMLSS-Modelle (Generalized Additive Models for Location, Scale and Shape) verwendet werden. Speziell für Daten mit Autokorrelation können Modelle aus der Zeitreihentheorie oder Geostatistik übernommen werden.

### 5.1.1. Lineare Modelle

Ein lineares Modell folgt der Modellgleichung  $\hat{y} = \beta_0 + \beta_1x_1 + \dots + \beta_nx_n$ , wobei in  $\mathbf{X}$  (gegebenenfalls transformierte) Kovariablen (keine Basisfunktionen) enthalten sind. Es gilt die strikte Normalverteilungs- und Unabhängigkeitsannahme für die Residuen. Werden dabei die kontinuierlichen Versionen der Kovariablen verwendet, so wird damit eine lineare Abhängigkeit implizit angenommen. Dies kann durch die Verwendung der diskretisierten Version umgangen werden, beziehungsweise die Gültigkeit dieser Annahme kann damit überprüft werden. Veränderungen des Einflusses einer Kovariablen in Abhängigkeit einer oder mehrerer anderer Kovariablen werden als Interaktionen der Form  $\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}(x_1 \cdot x_2)$  in die Modelle mit aufgenommen.

Zur Schätzung des Modells wird meist die Methode der kleinsten Quadrate verwendet, wobei die Quadratsumme der Residuen  $S = \|y - \hat{y}\|^2$  minimiert wird.

Ist das Modell geschätzt, können die einzelnen Koeffizienten  $\beta_i$  jeweils mit einem t-Test auf Verschiedenheit von Null getestet werden. Ist das Ergebnis eines dieser Tests nicht signifikant, bedeutet das, dass der zugehörige Koeffizient so ungenau geschätzt wurde, dass sein Konfidenzintervall die Null enthält und ihm wird im Allgemeinen keine große erklärende Wirkung zugeschrieben. Wird beispielsweise die Auswirkung von Nährstoffgehalt des Bodens und Höhe der  $\text{CO}_2$ -Konzentration auf die Wachstumsgeschwindigkeit von Pflanzen modelliert und der t-Test des Koeffizienten der  $\text{CO}_2$ -Konzentration wäre nicht signifikant, so würde man der Variablen  $\text{CO}_2$  eine geringe Erklärungskraft zuschreiben und überprüfen, ob sie sogar komplett aus dem Modell genommen werden kann. Andererseits muss dabei immer bedacht werden, dass die Signifikanz des t-Tests von der Stichprobengröße abhängt und somit am Besten auch die Größe des Koeffizienten an sich in die Interpretation mit einfließen sollte. Die Erklärungskraft des gesamte Modell, also aller  $\beta_i$  simultan, kann mit einem F-Test auf Verschiedenheit von einem Modell mit nur einem Interzept  $\beta_0$  getestet werden.

Die Modellgüte wird über Diagnoseplots für die Residuen und mit dem Anteil der vom Modell erklärten Varianz  $R^2$  beurteilt. Als Diagnoseplots werden Q-Q-Plot verwendet, welche die Verteilung der Residuen gegen eine Normalverteilung auftragen. Dabei entspricht die  $y = x$ -Diagonale einer Übereinstimmung der Verteilungen und jede Abweichung von dieser Geraden weist auf eine Abweichung der Verteilung hin. Weiterhin werden die vom Modell vorhergesagten Werte gegen die tatsächlichen Werte aufgetragen. Streuen diese Punkte nicht gleichmäßig um eine Konstante, so ist die Unabhängigkeitsannahme oder Verteilungsannahme verletzt. Weisen diese Punkte einen Trend auf, so sind nicht alle Abhängigkeiten im Modell passend modelliert. Sind die Voraussetzungen wie Normalverteilung der Residuen nicht hinreichend gut erfüllt, kann das Ergebnis des t-Tests der Koeffizienten verfälscht werden und eine Verzerrung des Betrags der Koeffizienten auftreten.

Möchte man mehrere Modelle miteinander vergleichen, so bietet sich zuerst eine Beurteilung über die Modellgüte an. Zusätzlich können die Modelle per ANOVA, Likelihood-Ratio-Test oder dem AIC- und BIC-Kriterium, welches die Komplexität der Modelle in der Beurteilung berücksichtigen, verglichen werden.

Generalisierte lineare Modelle erlauben die Anwendung auf nicht-Normalverteilte Daten und die Verwendung einer etwaigen (monotonen) Link-Funktion  $g$  auf die Zielvariable  $y$ . Die Modellgleichung lautet  $g(y) = \hat{y} + \varepsilon$ , wobei  $Y$  jetzt einer Verteilung aus der Exponentialfamilie folgen darf. Der Schätzalgorithmus beruht hierbei auf einer iterativen Strategie, was bedingt, dass die Tests (zum Beispiel t-Test für Koeffizienten, Modellvergleiche mit ANOVA) nunmehr approximativ aus dem zentralen Grenzwertsatz folgen.

### 5.1.2. (Generalisierte) additive Modelle

Additive Modelle bieten die Möglichkeit, Nicht-Linearitäten im Zusammenhang zwischen Ziel- und Kovariablen stetig, also ohne Diskretisierung einer eigentlich stetigen Kovariable, zu modellieren. Wenn also beispielsweise der Koeffizient der Schiffsemissionen mit der Größe der Schiffsemission variiert, das heißt,  $\beta_{\text{Schiff}}$  in der Modellgleichung  $\hat{y} = \beta_0 + \beta_{\text{Schiff}} \cdot x_{\text{Schiffsemission}}$  eigentlich ein  $\beta_{\text{Schiff}}(x_{\text{Schiffsemission}})$  ist, so kann dies im linearen Modell dadurch beschrieben werden, dass man die Schiffsemissionsvariable diskretisiert, also ihren Wertebereich in Abschnitte auf-

teilt und für diese jeweils einen eigenen Koeffizienten schätzt. Die Problematik an diesem Vorgehen ist die Wahl der Schnittpunkte für die Aufteilung, welche bei Verwendung von additiven Modellen durch die Wahl eines Glattheitsparameters ersetzt wird. Diese Wahl kann jedoch algorithmisch optimiert werden, im Gegensatz zur Wahl der Schnittpunkte. Bei den im Rahmen dieser Arbeit verwendeten kubischen Regressionssplines als Basisfunktionen wird die Glattheit über die Anzahl der Schnittpunkte (auch Knotenpunkte) definiert.

Der Unterschied zwischen (generalisierten) linearen und additiven Modellen liegt somit in der Art und Weise, wie die Kovariablen ins Modell aufgenommen werden. Während in den linearen Modellen die Funktionen  $f$ , welche über  $\hat{y} = \beta_0 + \beta_1 f(x_1)$  auf die Kovariablen angewendet werden können, lediglich eine Transformation, wie zum Beispiel Logarithmus oder Quadrat, beinhalten können, ist die einzige Voraussetzung an die Funktionen  $f$  im additiven Modell, dass sie glatt sein müssen. Die Modellgleichung lautet also  $g(Y) = \beta_0 + \beta_{i1}f_{i1}(x_{i1}) + \beta_{i2}f_{i2}(x_{i2}) + \dots + \beta_{j1}f_{j1}(x_{j1}) + \beta_{j2}f_{j2}(x_{j2}) + \dots + \varepsilon_i$ , wobei  $i$  und  $j$  hier verschiedene Kovariablen kennzeichnen. Die glatten Funktionen  $f$  werden dabei abschnittsweise aus Überlagerung von Basisfunktionen an die Daten angepasst. Auch zwei-, drei- oder  $n$ -dimensionale Abhängigkeiten, also Interaktionen zwischen zwei Kovariablen, können über die Verwendung der entsprechend  $n$ -dimensionalen Basisfunktionen ins Modell aufgenommen werden. Die Wahl der Basis findet vor der Modellberechnung statt, die Glattheit der Funktion wird (zwar durch Vorgaben beschränkbar) während der Modellberechnung mit geschätzt, also optimiert. Auch hier ist der Schätzalgorithmus iterativ, woraus nur eine approximative Gültigkeit der Tests folgt. Die Verteilungsannahmen sind die gleichen wie im GLM.

### 5.1.3. Geostatistik

Eine etwas andere Betrachtungsweise kommt aus dem Gebiet der Geostatistik. Eine ausführliche Beschreibung, auch für die Verbindung von Geostatistik und Raum-Zeit-Daten, ist in Cressie und Wikle (2011) [16] enthalten. Hier wird die Korrelation innerhalb der Daten  $Y$  durch Festlegung einer Kovarianzfunktion berücksichtigt. Die Kovarianzfunktion wird dabei so gewählt, dass sie das empirische Variogramm (Erklärung in Kapitel 2.2.2.3) interpoliert. Dabei wird davon ausgegangen, dass die Daten einem Gauß'schen Prozess entstammen. Häufig verwendete Kovarianzfunktionen sind sphärische, exponentielle oder Matérn-Funktionen. Diese Methodik heißt nach ihrem ersten Anwender D. G. Krige auch Kriging. Wird dabei zusätzlich ein linearer Trend in den Daten mit modelliert, spricht man von universellem Kriging oder auch Kriging mit Drift. Im Kontext von Raum-Zeit-Daten kann die Kovarianzfunktion tatsächlich auch dreidimensional (zwei Raum- und eine Zeitdimension) angesetzt werden. Alternativ kann ein rein räumliches Modell pro Zeitpunkt berechnet werden.

### 5.1.4. Zeitreihenmodelle

Es gibt viele Anwendungen, in denen eine Abhängigkeit von der Zeit modelliert werden muss. Im Fall von Raum-Zeit-Daten kann pro zwei- beziehungsweise dreidimensionalem Pixel (also pro Koordinaten-Punkt im Raum) ein Modell für die Abhängigkeit von der Zeit angesetzt werden. Im Rahmen dieser Arbeit beschrän-

ken sich diese Modelle auf Regressionsmodelle, bei denen die Kovariablen zusätzlich um ein oder mehrere Zeitschritte verschoben werden. Es können jedoch bei Bedarf auch Modelle, wie zum Beispiel ARMA oder ARIMA, aus der Zeitreihentheorie übernommen werden. Eine Einführung im Kontext von Raum-Zeit-Daten bietet [16], im Kontext von meteorologischen Daten [69].

### 5.1.5. Modellgütekriterien

Zum Vergleich von Modellen werden verschiedene Kriterien verwendet, die jeweils unterschiedliche Aspekte des Modells beurteilen. Am einfachsten zu interpretieren ist  $R^2$ , welches den Anteil der vom Modell erklärten Varianz an der Gesamtvarianz der Daten beschreibt. Allgemein wird ab einem Wert von 70-80% für  $R^2$  von einem guten Modell gesprochen, wobei dies teilweise auch von der Konvention des jeweiligen Wissenschaftsgebiets abhängt. Ein Nachteil bei der Verwendung von  $R^2$  ist, dass es alleine durch die Komplexität des Modells steigt und für viele komplexere Modelle, vor allem hierarchisch aufgebaute Modelle, nicht eindeutig zu berechnen ist. Als Alternative bieten sich das Akaike-Informationskriterium (AIC) oder das Bayes-Schwarz-Informationskriterium (BIC) an. Von diesen beiden 'bestraft' das BIC komplexere Modelle stärker. Beide berechnen sich mit der Formel  $-2 \cdot \loglik + k \cdot npar$ , wobei  $\loglik$  die Log-Likelihood des Modells,  $npar$  die Anzahl der Parameter des Modells und  $k = 2$  (AIC) bzw.  $k = \log(n)$  (BIC) der Strafterm für die Komplexität ist.  $n$  ist dabei die Anzahl der im Modell verwendeten Datenpunkte. Aus der Definition folgt, dass ein kleineres AIC und BIC für ein besseres Modell steht, wobei anders als bei  $R^2$  der absolute Wert keine Aussagekraft hat.

## 5.2. Regressionsmodelle für den Beispieldatensatz

Die untersuchte Fragestellung 'Ist die Differenz der betrachteten Variablen zwischen den beiden Simulationen signifikant?' kann im Kontext der Regressionsmodellierung mit einem F-Test der Koeffizienten des jeweiligen Modells beantwortet werden. Ebenso kann die Fragestellung in 'Kann die Differenz durch eine Abhängigkeit von den Eigenschaften  $x_i$  der Daten beschrieben werden?' umformuliert werden. Damit kann im einzelnen beurteilt werden, ob die Differenz bestimmte Muster in Raum und Zeit oder in Abhängigkeit der Emissionen und des Niederschlags aufweist. Finden sich keine solchen Muster beziehungsweise nur welche, die keine Signifikanz aufweisen, wird das als 'kein Unterschied zwischen den Simulationen' interpretiert.

Es werden zwei verschiedene Situationen untersucht. Im ersten Fall wird davon ausgegangen, dass keine weiteren Informationen, außer den Raum- und Zeitkoordinaten zur Verfügung stehen. Im zweiten Schritt wird untersucht, ob sich das Modell durch Hinzunahme der Informationen über die Schiffsemissionen und den Niederschlag verbessern lässt. Die Modelle werden nicht nur nach den Modellsелеktionskriterien AIC, BIC und  $R^2$  beurteilt, sondern auch im Hinblick auf die Verteilung ihrer Residuen. Dem Prinzip von Occam's Rasiermesser folgend werden die Modelle ausgehend von dem einfachsten zu immer weiter zunehmender Komplexität aufgebaut.

Die Wahl des Modells wird am Beispiel der Differenz der Sulfat-Konzentration für die unterste Höhenschicht durchgeführt, das Verfahren für die anderen Varia-

blen ist analog. Die Auswertung der besten Modelle findet sich abschließend in Kapitel 5.2.5.

Zur Erinnerung sind hier noch einmal kurz die wesentlichen Eigenschaften der Konzentrationsdifferenz der Sulfat-Aerosole beschrieben. Es ist eine zeitliche und räumliche Korrelation in den Daten vorhanden, welche mit zunehmender Entfernung abnimmt und räumlich weitgehend isotrop ist. Allerdings kann diese Autokorrelation getrennt berücksichtigt werden (Ergebnis der deskriptiven Analyse in Kapitel 2.2.2.3). Andererseits zeigt die Differenz zusätzlich einen Jahresgang, der räumlich gesehen unterschiedlich stark ausgeprägt ist. Insgesamt gesehen ist die Konzentrationsdifferenz nicht normalverteilt und ihre Zahlenwerte sind, numerisch gesehen, relativ klein (Größenordnung  $10^{-9}$  bis  $10^{-14}$ , was eventuell zu Instabilitäten der Algorithmen führen kann. Die Kovariablen Schiffsemission und Landemission sind in verschiedenen Varianten vorhanden: unverändert, standardisiert, diskretisiert durch Runden der standardisierten Variante, diskretisiert durch Einteilung in Quantile und mit sechs verschiedenen räumlichen Glättungen. Der Niederschlag ist ebenfalls in diesen Versionen vorhanden, bis auf die Glättung.

## 5.2.1. Lineare Modelle

### 5.2.1.1. Lineare Modelle ohne Aufteilung nach Raum oder Zeit

Es werden Modelle der Form  $Konzentration = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon$  berechnet, wobei  $x_i$  für die Kovariablen, hier Schiffsemissionen, Landemissionen, Niederschlag (Abbildungen 2.2 und A.1), Breitengrad, Längengrad und Zeit (in Form von Monaten und/oder Jahren oder als Index für die einzelnen Zeitpunkte), sowie etwaige Interaktionen derselben steht.

Mit Modellen, in welchen nur die Koordinaten, also Längengrad, Breitengrad und Zeit, als kontinuierliche Kovariablen enthalten sind, kann eine maximale Modellgüte von  $R^2 \approx 27\%$  erreicht werden. Eine ähnlich große Modellgüte von  $R^2 \approx 37\%$  erreichen auch die Modelle mit den Emissions- und Niederschlagsmustern als Kovariablen, wobei dabei hauptsächlich die Schiffsemissionen zur Modellgüte beitragen. Man kann die Modellgüte weiter steigern, wenn statt den kontinuierlichen Kovariablen die diskreten Versionen verwendet werden. Auch bei Verwendung des geglätteten Schiffsmusters statt des ursprünglichen, steigt  $R^2$  (auf ungefähr 46%). Nimmt man alle möglichen Variablen mit auf, wird die Modellgüte noch einmal leicht gesteigert und mit der Kombination aus Schiffsemissionen und diskreten Landemissionen, Niederschlag und Zeitpunkten erreicht man die maximale Modellgüte aller untersuchten linearen Modelle von 49%. Die entsprechenden Werte können in Tabelle C.4.1 nachgeschlagen werden. Es zeigte sich bei der Betrachtung der Koeffizienten der diskreten Kovariablen, dass eigentlich keine Linearität über dem Wertebereich der jeweiligen Kovariable angenommen werden kann. Die Berechnung von Modellen mit diskreten Kovariablen, welche eine sehr feine Aufteilung des Wertebereiches aufweisen, oder mit Interaktionen von diskreten Kovariablen, war allerdings aufgrund des hohen Ressourcenverbrauchs dieser Modelle nicht möglich.

Betrachtet man die Residuenverteilung in der Abbildung C.4.2 des besten Modells, so stellt sich heraus, dass die Residuen nicht normalverteilt sind und sie einen leichten Trend in Bezug auf die vorhergesagten Werte aufweisen. Die Koeffizienten, nachzuschlagen in der Tabelle C.4.2, sind alle zum Niveau  $\alpha = 0.05$

signifikant, wobei dies natürlich auch mit der großen Stichprobe zusammenhängt. Da durch die Standardisierung und Diskretisierung die Koeffizienten nicht mehr mit physikalischen Einheiten behaftet sind, können ihre Größenordnungen direkt miteinander verglichen werden. Man sieht, dass sie sich alle in der gleichen Größenordnung von  $10^{-11}$  befinden und damit einen ungefähr gleich großen Einfluss der Kovariablen anzeigen. In Übereinstimmung mit der physikalischen Interpretation, dass der Niederschlag Aerosole aus der Luft auswäscht, sind die Koeffizienten desselben negativ, und die der Emissionen positiv. Die Koeffizienten für die einzelnen Zeitpunkte zeigen einen klaren Jahresgang mit höheren Werten im Sommer und niedrigeren im Winter. Hier ist offensichtlich die zeitliche Variabilität nicht vollständig durch die Zeit-Kovariablen erklärt, aber Modelle, die einen Jahresgang, also nur die Monatszahl oder die Monatszahl und das Jahr als Kovariablen enthalten, wurden von den Modellgütekriterien als leicht schlechter beurteilt.

#### **5.2.1.2. Generalisierte Lineare Modelle und GAMLSS**

Es gibt die Möglichkeit, eine andere Verteilungsannahme zu treffen, wodurch sich die Residuenverteilung verbessern könnte. Allerdings führen die für GLMs üblichen, hier in Frage kommenden Verteilungen nicht zu einer offensichtlichen Verbesserung (Diagnoseplots in Abbildung C.4.3). Berechnet man ein GAMLSS, können weitere Verteilungen ausprobiert werden. Manche davon, zum Beispiel Johnson's SU-Verteilung oder die sinh-arcsinh-Verteilung, führen tatsächlich zu annähernd normalverteilten Residuen (Abbildung in C.4.3). Trotzdem ist noch eine ungleichmäßige Streuung der Residuen in den Diagnoseplots erkennbar. Aufgrund der wesentlich längeren Rechenzeit der GAMLSS-Modelle wurde dieser Ansatz nicht weiter verfolgt.

Es können jetzt zwei weitere Ansätze verfolgt werden. Entweder man versucht mit einer räumlichen Modellierung pro Zeitpunkt die räumliche Korrelation zu beschreiben oder mit einer zeitlichen Modellierung pro Pixel beziehungsweise Raumkoordinatenpunkt die zeitliche Korrelation. Da der Jahresgang relativ offensichtlich ist, wurde zuerst letzteres versucht.

#### **5.2.1.3. Lineare zeitliche Modelle pro Pixel**

In dieser Art von Modellen treten nicht nur die Kovariablen an sich auf, sondern auch die um einen 'Lag' versetzten Werte. Unter anderem kann auch eine verschobene Version der Variablen selber als Kovariable auftreten. Das Ziel der Modellierung ist, den Jahresgang adäquat zu beschreiben. Die untersuchten Modelle lassen sich in zwei Kategorien aufteilen. Kategorie 1 sind alle Modelle, bei welchen nur die zeitliche Autokorrelation, also der Zusammenhang der Differenz mit sich selber (aber zeitlich verschoben), modelliert wird. In die Modelle der Kategorie 2 wurden die (teilweise auch zeitlich verschobenen) Emissions- und Niederschlagsmuster als Kovariablen verwendet. Eine Auswahl an Diagnoseplots für verschiedene Kovariablenstrukturen und -kombinationen findet sich in C.4.4. Die folgende Auswertung bezieht sich auf die Abbildungen A.15, A.16, A.17 und C.4.4 (Punkt eins), steht aber exemplarisch für die weiteren Modelle aus C.4.4.

Es wird jeweils ein Modell pro Kategorie für die Differenz der Sulfat-Konzentration in der untersten Höhengschicht beschrieben. Das Beispielmmodell für Kategorie 1

enthält als Kovariablen die um einen und sechs Monate verschobene Differenz selbst. Die räumliche Verteilung der Koeffizienten ist in Abbildung 5.1 zu sehen. Der Interzept ist über den Ozeanflächen großräumig homogen, während über den Landflächen eine starke kleinskalige Variabilität zu sehen ist. Auch der Koeffizient für die direkte Autokorrelation von einem Monat zeigt dieses Verhalten. Während der Interzept weitgehend positiv ist, ist der Lag-1-Koeffizient größtenteils negativ. Da beide in der gleichen Größenordnung von  $10^{-1}$  liegen, lässt sich vermuten, dass eigentlich beide zusammen eine Art Interzept bilden. Interessant ist die räumliche Verteilung des Koeffizienten für die Halbjahres-Verschiebung, denn diese zeigt große Ähnlichkeit mit dem Muster der Schiffsemissionen. Die Größenordnung des Koeffizienten liegt im Bereich der Differenz selber. Die Diagnoseplots zeigen teilweise eine große Abweichung von der Normalverteilungsannahme und verbleibende Autokorrelation in den Residuen. Die Modellgüte  $R^2$  liegt nur im Bereich des Nordpazifik und Nordatlantik ( $\approx 80\%$ ) über 30%.

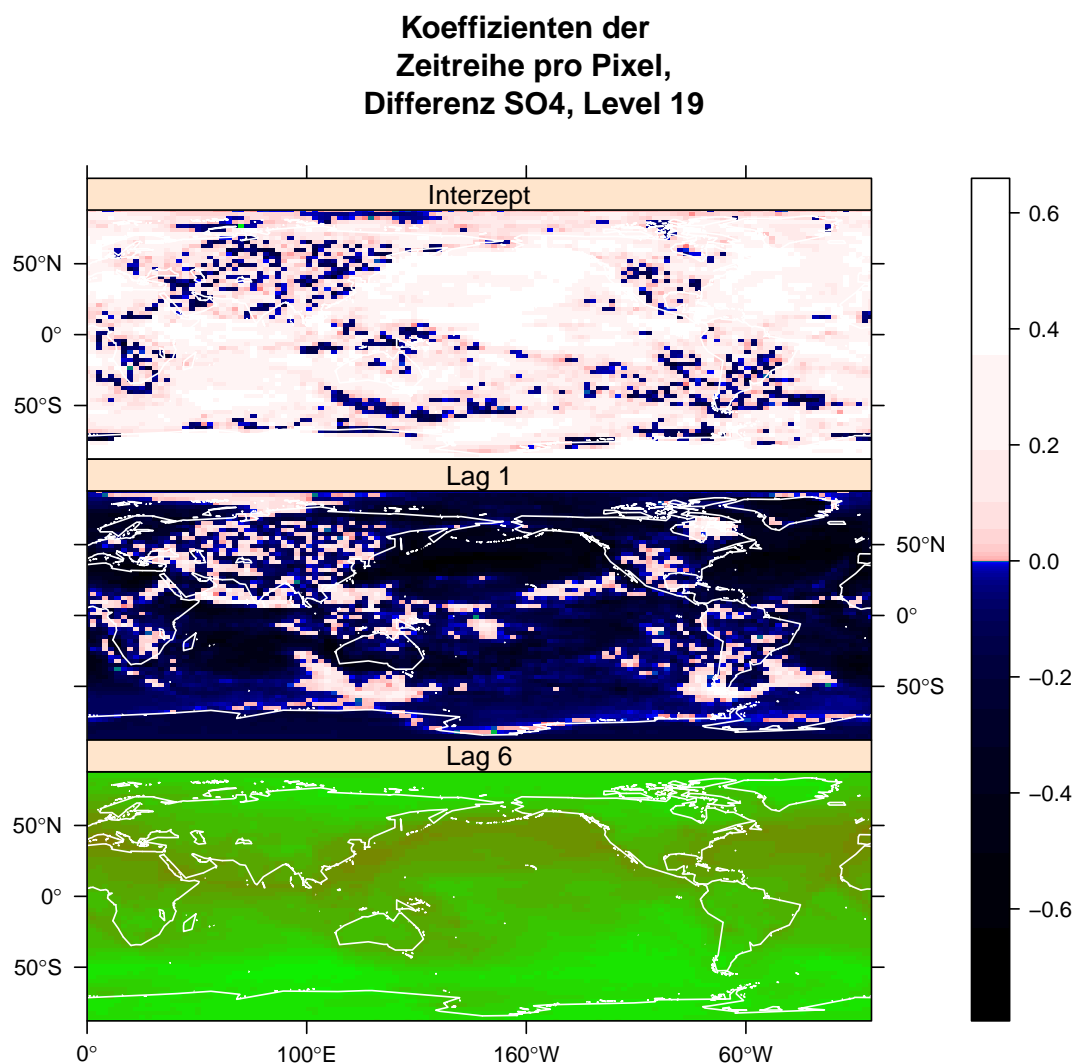


Abb. 5.1.: Räumliche Verteilung der Koeffizienten der linearen Modelle pro Pixel für die Sulfat-Differenz der untersten Schicht mit dem um einen und sechs Monate Wert der Differenz selbst als Kovariablen



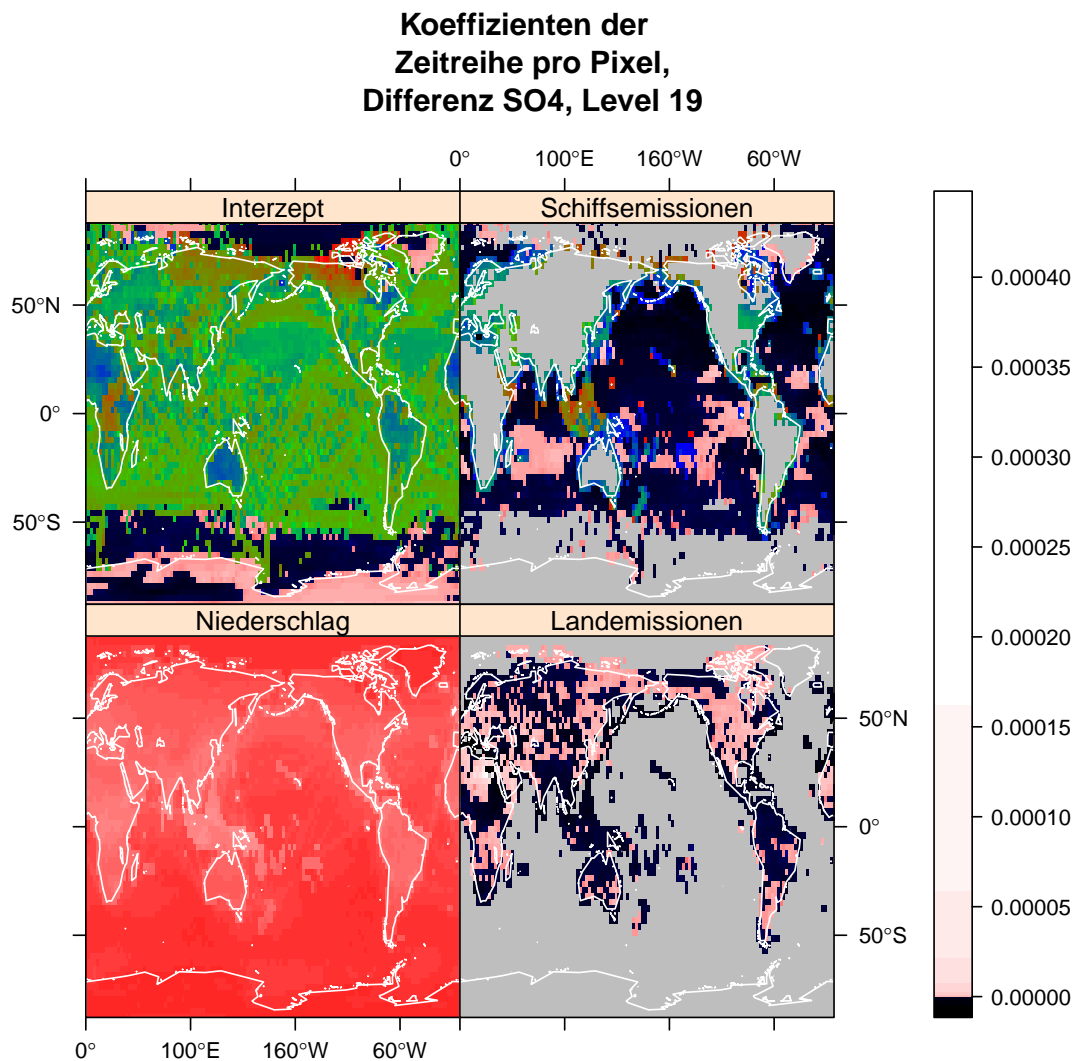


Abb. 5.2.: Räumliche Verteilung der Koeffizienten der linearen Modelle pro Pixel für die Sulfat-Differenz der untersten Schicht mit den Emissionen und dem Niederschlag als Kovariablen

Als Beispiel für ein Modell der Kategorie 2 wurden die unveränderten Emissions- und Niederschlagsvariablen als Kovariablen verwendet. Die räumliche Verteilung der Koeffizienten ist in Abbildung 5.2 dargestellt. Bei diesem Modell zeigt der Interzept über den Ozeanen eine räumlich homogene kleinskalige Variabilität um die Null. Über den Kontinenten der Südhalbkugel ist er leicht negativ, über der Arktis und vor der Küste der Antarktis stark negativ und über der Antarktis selber, Nordkanada und Grönland stark positiv. Die Koeffizienten für die Schiffsemissionen sind im Bereich des Äquators positiv und über dem Nordatlantik und -pazifik negativ. Dabei zeigen sie keine kleinskalige Variation und tendieren an den Küsten gegen Null. Die Koeffizienten der Landemissionen variieren räumlich viel kleinskaliger. Ausgeprägte Maxima (positiv) befinden sich über Nordamerika, Korea und der Sahara, Minima (negativ) vor allem in Küstennähe. Die durchwegs positiven Koeffizienten für den Niederschlag sind um gute zwei Größenordnungen kleiner als der Interzept und die der Schiffs- und Landemissionen, welche sich unge-

fähr in der gleichen Größenordnung bewegen. Die Niederschlagskoeffizienten sind räumlich sehr homogen verteilt mit einer klaren Land-Ozean-Unterscheidung. Maxima befinden sich in Küstennähe, südlich der Sahara und über Ozeanien. Minimal werden die Koeffizienten über den Polen. Dies führt zu dem Schluss, dass die Niederschlagskoeffizienten bei den Küstenpixeln als eine Art Ausgleich zwischen den beiden Emissionskoeffizienten dienen. Die Modellgüte hat sich im Vergleich zum Beispielmodell der Kategorie 1 nicht stark verändert. Es zeigt sich immer noch ein starker Jahresgang in den Residuen und eine nicht unerhebliche räumliche Korrelation.

Im Vergleich zu den rein linearen Modellen hat sich hier also keine wesentliche Verbesserung ergeben, eher sogar eine geringere gesamte Modellgüte.

#### **5.2.1.4. Lineare Modelle pro Zeitpunkt**

Als nächstes werden Modelle für jeden Zeitpunkt, also jeden Monat, einzeln betrachtet, um den Fokus auf die richtige Modellierung der räumlichen Korrelation zu legen, wodurch die zeitliche Korrelation beziehungsweise Schwankung in die Koeffizienten verschoben wird.

Es wurden drei verschiedene Modelle untersucht. Modell 1 enthält nur die Koordinaten als diskrete Kovariablen. Modell 2 enthält nur die Emissions- und Niederschlagsmuster in der vom linearen Modell als am besten vorgeschlagenen Glättung oder Diskretisierung (ablesbar aus Tabelle C.4.2). Modell 3 enthält auch nur die Emissionen und den Niederschlag, aber in der diskreten Variante mit Rundung des standardisierten Musters. Bei Modell 1 (Abbildung 5.3) mit den Koordinaten als Kovariablen zeigt der Breitengrad vor allem in Äquatornähe und nördlich davon einen klaren Jahresgang, der Längengrad eher durchgehend bis auf den Bereich um den Greenwich-Meridian. Dabei sind die Koeffizienten für die Breitengrade überwiegend positiv, die der Längengrade negativ. Auch der Interzept schwankt relativ gleichmäßig im eher positiven Bereich. Dabei bewegen sich die Größenordnungen der Koeffizienten im gleichen Bereich. Die Modellgüte liegt mit unregelmäßigen Veränderungen um circa 30%. Modell 3 mit den diskreten Emissions- und Niederschlagswerten zeigt eine wesentlich stärker variable Modellgüte zwischen 20-60%, welche regelmäßig für die Sommermonate ein Maximum und in den Wintermonaten das Minimum erreicht. Der Interzept bewegt sich mit relativ gleichmäßigen Schwankungen im positiven Bereich. Auch die Koeffizienten der Schiffsemissionen sind durchgehend positiv und zeigen einen klaren Jahresgang mit Sommermaxima und Winterminima. Dabei ist die Schwankungsbreite umso höher, je größer die Emissionskonzentration ist. Dies ist ein klarer Hinweis auf ein zeitlich nicht-lineares Verhalten, da die Winterminima der Koeffizienten nahe beieinander liegen, die Sommermaxima jedoch weit auseinander. Ein ähnliches Verhalten zeigen die Koeffizienten des Niederschlags, nur dass diese um die Null schwanken, genauso wie die Koeffizienten der Landemission. Bei diesen ist jedoch kein Jahresgang erkennbar. Eine leicht höhere Modellgüte von 70% im Sommer erreicht man bei Modell 2 (Abbildung in C.4.5).

#### **5.2.2. Geostatistik (räumliche Modelle pro Zeitpunkt)**

Mit der Methodik des Kriging lässt sich über die Verwendung einer Kovarianzfunktion die räumliche Korrelation mit in ein lineares Modell aufnehmen. Da ein

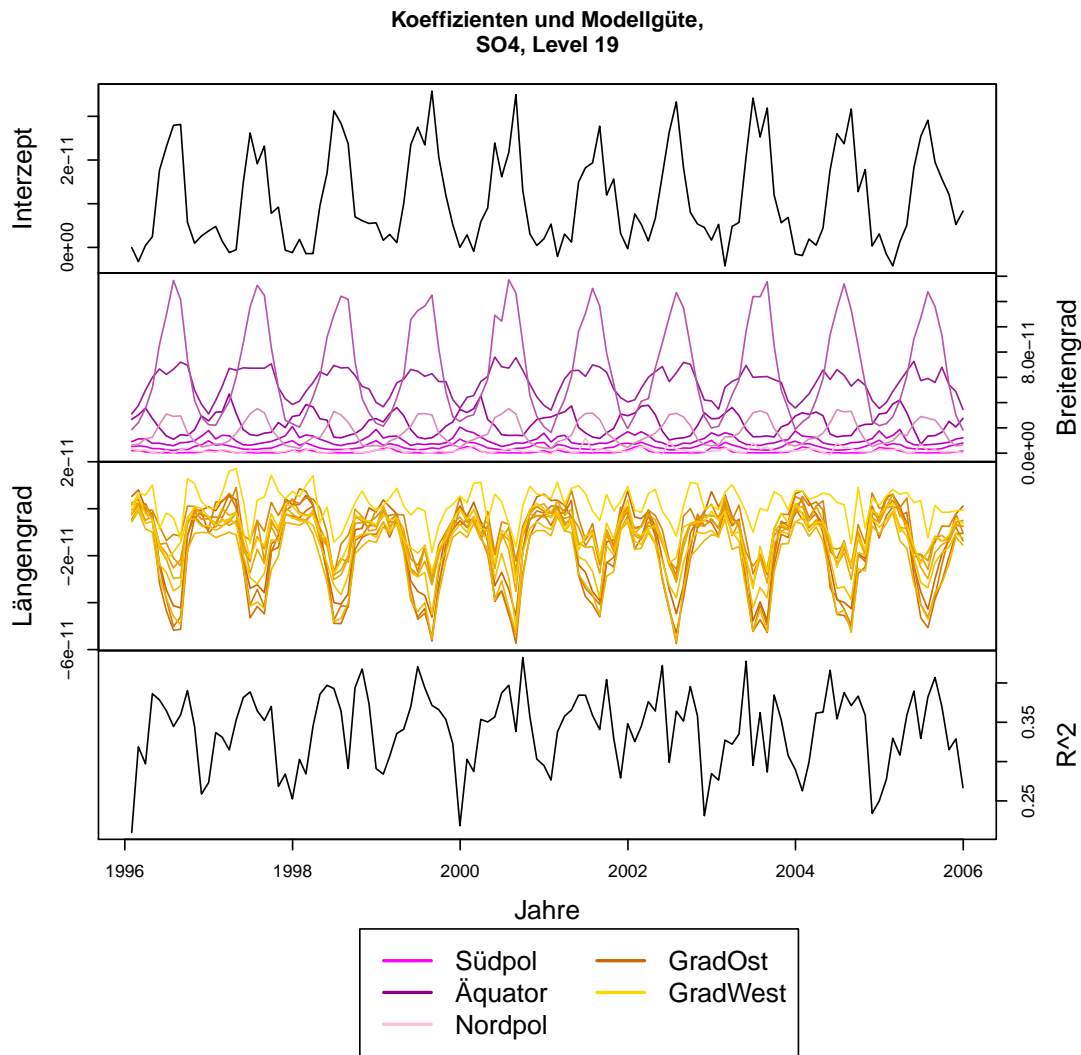


Abb. 5.3.: Zeitlicher Verlauf der Koeffizienten des Modells 1 der linearen Regression pro Zeitpunkt für die Sulfat-Konzentration der untersten Höhengschicht

vollständiges Raum-Zeit-Modell (vergleiche Abschnitt 5.1.3) aufgrund der RAM-Speicherbegrenzung nicht berechnet werden konnte, wurden nur für die einzelnen Zeitpunkte Modelle berechnet. Die Wahl der Art und Schätzung der Parameter der Kovarianzfunktion wurde in diesem Beispiel dem automatischen Algorithmus der *autoKrige*-Funktion aus dem R-Paket *automap* [29] überlassen. Es werden beispielhaft zwei Modelle mit verschiedenen linearen Trends betrachtet. Modell 1 enthält keinen Trend, besteht also nur aus der räumlichen Interpolation (ordinary Kriging). In Modell 2 wird die lineare Abhängigkeit von den Emissions- und Niederschlagsmustern mit berücksichtigt (universal Kriging). Die entsprechenden Koeffizienten- und Diagnoseplots sind in C.4.6 zu sehen.

Beide Modelle weisen räumlich wesentlich unkorrelierte Residuen auf als die oben betrachteten linearen Modelle. Zwar ist bei Modell 1 eine leichte großskalige Ähnlichkeit (vor allem im Bereich der Nordhemisphäre) erkennbar, in Modell 2 ist diese auf den Bereich über den Ostküsten Chinas und Nordamerikas und dem Mit-

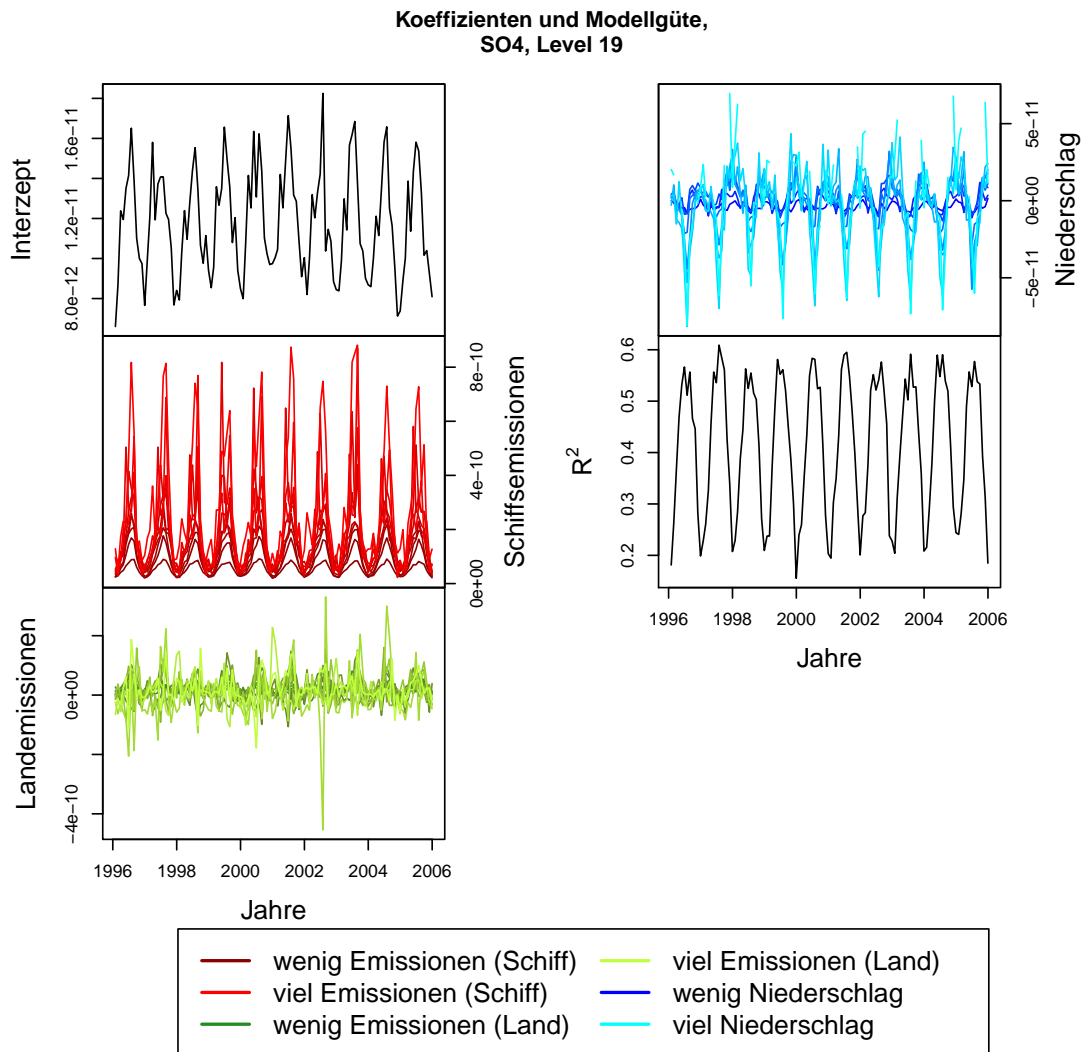


Abb. 5.4.: Zeitlicher Verlauf der Koeffizienten des Modells 3 der linearen Regression pro Zeitpunkt für die Sulfat-Konzentration der untersten Höhengschicht. Dabei ist die Helligkeit der Farben invers zur Stärke der Emissionen beziehungsweise des Niederschlags

telmeer begrenzt. Der Unterschied zwischen den Koeffizienten des Kriging-Modells und des linearen Modells ist verschwindend gering. Ein Nachteil des Kriging ist allerdings, dass nur lineare Abhängigkeiten von den Kovariablen berücksichtigt werden können.

### 5.2.3. additive Modelle

Der große Vorteil der Verwendung von additiven Modelle ist die große Anzahl von Möglichkeiten ein Modell aufzustellen, was meist eine gute Anpassung an die Daten ermöglicht. Dies bedeutet aber auch, dass viele Modelle berechnet werden müssen, aus welchen dann das beste ausgewählt werden kann. Eine Auflistung aller berechneten Modelle mit ihrer jeweiligen Modellgüte (per AIC, BIC und  $R^2$ ) ist

Kovariablen	Koordinaten, Emissions- und Niederschlagsmuster, beide zusammen
Struktur der Kovariablen	standardisiert, geglättet (mit Filtern)
Art der Basis-Funktionen	kubische Regressionssplines, zyklische kubische Regressionssplines, Gauß'scher Prozess (Kriging)
Beschränkung für den Glattheitsparameter <sup>1</sup>	niedrige Auflösung (5 Knoten), mittlere Auflösung (10 Knoten), hohe Auflösung (20 Knoten)
Interaktionen zwischen den Variablen	keine, räumliche (mit Längen- und Breitengrad), zeitliche (mit Monaten und Jahren), raum-zeitliche (mit den Koordinaten insgesamt)

Tabelle 5.1.: Auflistung der verschiedenen verwendeten Spezifizierungsmöglichkeiten bei der additiven Modellierung

in Tabelle von C.4.1 enthalten. Dabei gibt die erste Spalte die Modellnummer an, unter welcher auch die zugehörigen Koeffizienten- und Diagnoseplots in C.4.7 zu finden sind. Dabei ist zu beachten, dass in den Koeffizientenplots die Koeffizienten nicht direkt gegen die Werte der Variablen, sondern das jeweilige Produkt aus Koeffizient und Variablenwert gegen den Wert der Variablen aufgetragen ist.

Zur Ressourcen-sparenden Berechnung wurde die Funktion *bam* aus dem R-Paket *mgcv* [72] verwendet. Es wurden verschiedene Kombinationen von Kovariablen und ihren Strukturen untersucht, welche kurz in Tabelle 5.1 aufgelistet sind.

Zur Auswahl des besten Modells unter den berechneten werden die Modelle mit dem AIC und dem BIC verglichen, wobei das BIC einfachere Modelle (die wesentlich weniger Parameter enthalten) stärker bevorzugt. In diesem Fall stimmen die Beurteilungen von AIC und BIC überein und das so gewählte Modell erklärt einen Devianzanteil von ungefähr 70%. Es enthält sowohl alle Variablen (Koordinaten mit der Zeit als Monate und Emissions-/Niederschlagsmuster) als auch sämtliche Interaktionen von jeweils drei der Variablen. Die zweit- und drittbesten Modelle enthalten jeweils alle Gauß- und Mittelwerts-geglätteten Schiffsmuster gleichzeitig, die Zeit ebenfalls als Monate, raum-zeitliche Interaktionen (inklusive der Interaktion zwischen Längengrad, Breitengrad und Monaten), Interaktionen zwischen den Emissionsmustern und der Zeit sowie die Interaktion von Längengrad und Breitengrad in hoher Auflösung (20 Knotenpunkte, Definition in 5.1.2).

Die Koeffizienten der Schiffs- und Landemissionen im besten Modell, zu sehen in Abbildung 5.5, zeigen einen klaren nicht-linearen positiven Anstieg, das heißt, je größer die Emissionen, desto höher die Differenz der Sulfat-Konzentration in der untersten Höschicht. Die Koeffizienten des Niederschlags hingegen zeigen eine klare Abnahme mit der Menge des Niederschlags an sich, vor allem für geringe Mengen. Für höhere Niederschlagsmengen vergrößert sich die Unsicherheit in der Schätzung, was vermutlich an der geringen Anzahl an Beobachtungen für diesen Bereich liegt. Der Verlauf der Koeffizienten für den Monat zeigt zwei Maxima, im April und im September, wobei er nur für das Septembermaximum positive Werte erreicht. Die Koeffizienten für den Breitengrad weisen für die Südhalbkugel

<sup>1</sup>Durch Festlegung der maximalen Anzahl an Knotenpunkten im Wertebereich der jeweiligen Kovariablen

ein lokales Minimum und für die Nordhalbkugel ein lokales Maximum auf, wobei die globalen Minima und Maxima jeweils am Nord- und Südpol liegen. Die Koeffizienten für den Längengrad haben zwei Maxima bei ungefähr  $0^\circ$  (Ostatlantik) und  $180^\circ$  (Pazifik) und zwei Minima dazwischen ( $100^\circ \approx$  Asien und  $270^\circ \approx$  Nordamerika). Dabei ist das Maximum über dem Atlantik und das Minimum über Nordamerika jeweils stärker ausgeprägt. Bei beiden Minima beziehungsweise Maxima weisen die Koeffizienten jeweils negative beziehungsweise positive Werte auf. Die Koeffizienten für die wichtigsten Interaktionen der jeweiligen Kovariablen sind in Abbildung , die restlichen in C.4.7 zu finden. Es zeigt sich für die Schiffs-, Landemissionen und den Längengrad eine Veränderung mit den Monaten. Die Koeffizienten der Emissionen sind im Sommer höher als im Winter. Die Koeffizienten des Längengrades zeigen ein Sommerminimum bei circa  $270^\circ$  und ein Sommermaximum bei  $0^\circ$ . Die Koeffizienten des Längengrades und des Niederschlages variieren nur sehr schwach im Jahresverlauf. Einzig in der Antarktis zeigt sich eine etwas größere Schwankung zwischen Herbst (Maximum) und Winter (Minimum). Die, wenn auch nur approximativ gültigen, p-Werte führen, wie in Tabelle A.2 zu sehen ist, für alle enthaltenen Kovariablen und Interaktionen zu einer eindeutigen Signifikanz (bei einem Test zum Niveau  $\alpha = 0.05$ ). Betrachtet man allerdings die Diagnoseplots für die Residuen in Abbildung C.4.7, so zeigt sich immer noch eine Nicht-Normalverteilung der Residuen und verbleibende räumliche und zeitliche Korrelationen.

#### 5.2.4. Vergleich der verwendeten Modellierungsstrategien

Es kann zwischen zwei Hauptstrategien unterschieden werden. Bei Strategie 1 wird das Modell auf den kompletten Datensatz (der untersten Schicht) angewendet (lineare Modelle, additive Modelle), wohingegen dieser bei Strategie 2 vorher nach Raum oder Zeit aufgeteilt wird und dann auf die einzelnen Teile (Pixel oder Zeitpunkte, also Monate) jeweils ein Modell angewendet (Zeitreihenmodelle, Kriging-Modelle). Bei diesen aufgeteilten Modellen war eine starke räumliche Schwankung (Zeitreihenmodelle) beziehungsweise starke zeitliche Schwankung (lineare Modelle pro Monat) der Modellgüte mit der jeweiligen Dimension, in welcher aufgeteilt wurde zu sehen ( $R^2$  zwischen 20% bis 70%). Dies lässt den Schluss zu, dass die Kovariablen abhängig von den Koordinaten in bestimmten Gebieten (vor allem Nordpazifik) und zu bestimmten Zeitpunkten (Sommer) eine größere erklärende Kraft haben, als bei den restlichen Raum- und Zeitpunkten. Der Vorteil der Strategie, nur ein Modell für den kompletten Datensatz anzusetzen, liegt in der Kompaktheit (nur ein Modell) der Darstellung und Ausgewogenheit der Berücksichtigung von Raum und Zeit. Deren Korrelation wird entweder nicht im Modell (lineare Modelle) berücksichtigt oder es sind beide im Modell (additive Modelle) enthalten. Die Modellgüte der linearen Modelle liegt unter der, welche durch alle anderen Ansätze mindestens erreicht wird. Die größte Modellgüte zeigen die additiven Modelle, was höchstwahrscheinlich durch ihre flexiblen Anpassungsmöglichkeiten an die Daten bedingt ist. Über alle betrachteten Modelle hinweg ist insgesamt zu beobachten, dass die Modelle mit den Emissions- und Niederschlagsmustern im Vergleich zu den Modellen mit nur den Koordinaten eine fast doppelt so hohes  $R^2$  aufweisen.

### 5.2.5. Auswertung der Modelle für alle Variablen

Für die anderen betrachteten Variablen außer der Sulfat-Konzentration wurden jeweils die einfachen linearen Modelle, die linearen Modelle pro Zeitpunkt und die additiven Modelle berechnet, da sich diese bei den Sulfat-Konzentrationen als besonders informativ und gut interpretierbar herausgestellt haben. Die Ergebnisse und Abbildungen der linearen Modelle finden sich in C.4.2 und C.4.5. Das unten genauer beschriebene Verhalten der Koeffizienten der additiven Modelle bezieht sich, falls nicht anders erwähnt, stets auf das von AIC und BIC als am besten ausgewählte Modell (Tabelle C.4.1).

#### 5.2.5.1. Differenz der Sulfat-Aerosole

Bereits bei Betrachtung der Koeffizienten der linearen Modelle pro Monat ist erkennbar, dass sich die zeitliche Variation in der Konzentrationsdifferenz besonders stark auf die Schiffsemissionen zurückführen lässt. Es ist ebenso die Tendenz erkennbar, dass die Modelle mit Zusatzinformationen durch die Emissions- und Niederschlagsmuster eine bessere Modellgüte aufweisen (bis zu 40% mehr). Betrachtet man die additiven Modelle, so wird der direkte Zusammenhang zwischen Konzentrationsdifferenz und Emissionsmuster bei Hinzunahme der räumlich geglätteten Emissionsmuster (beschrieben in Kapitel 2.2) als linearer Zusammenhang geschätzt. Grundsätzlich wird von den Modellen ein positiver Zusammenhang berechnet, also eine größere Differenz bei höheren Schiffsemissionen. Bei den Landemissionen ist die Stärke des Zusammenhangs, gemessen durch den Wert der Koeffizienten, etwas niedriger. Beim Niederschlag ergibt sich kein einheitliches Bild. Während die Richtung des Zusammenhangs mit dem Niederschlag (Wechsel zwischen positiven und negativen Koeffizienten) in den linearen Modellen je nach Modellspezifikation wechselt, schätzen die additiven Modelle größtenteils einen negativen Zusammenhang, also eine sinkende Differenz bei mehr Niederschlag. Dies entspricht der physikalischen Interpretation, dass durch den Regen die Aerosole vermehrt ausgewaschen werden und durch den Schiffsverkehr mehr Aerosole in die Luft gelangen. Dabei ist der Einfluss des Schiffsverkehrs in den Sommermonaten größer als in den Wintermonaten. Werden alleine die Koordinaten als Kovariablen verwendet, so lassen sich Regionen mit großer geschätzter Differenz identifizieren. Großskalige Maxima liegen auf der Nordhalbkugel über dem Pazifik und dem Mittelmeer. Kleinskalig Maxima lassen sich vor allem die Industriegebiete über China und Europa identifizieren. Wird die kleinskalige Modellierung ohne Emissions- und Niederschlagsmuster durchgeführt, sind in den 2D-Koeffizienten des Längen- und Breitengrades die Haupt-Schifffahrtsrouten erkennbar (umso besser, je höher die räumliche Auflösung bei der Modellierung gewählt wurde).

Die Signifikanz der Koeffizienten ist in den linearen Modellen durchgängig gegeben (Tabelle C.4.2). Erst wenn die diskreten Versionen der Koordinaten oder Emissions-/Niederschlagsmuster verwendet werden, fällt der t-Test bei einigen Wertebereichen der Kovariablen nicht signifikant aus. Bei den additiven Modellen fällt außerdem auf, dass die 95%-Konfidenzbänder für die meisten Koeffizienten sehr schmal sind. Damit können die oben beschriebenen Trends mit hoher Sicherheit als von Null verschieden angenommen werden. Als Gesamtinterpretation bedeutet dies, dass hier ein wesentlicher Unterschied (in der Differenz) zwischen den beiden Simulationen zu finden ist, dessen räumliche Verteilung mit der räumlichen

Intensität des Schiffsverkehrs in Verbindung gebracht werden kann.

#### 5.2.5.2. Differenz der Nitrat-Aerosole

Auch hier zeigt sich in den linearen Modellen pro Monat mit den Emissions- und Niederschlagsmustern ein starker Jahresgang in den Koeffizienten für die diskretisierten Schiffsemissionen. Zusätzlich ist ein etwas weniger stark ausgeprägter Jahresgang in den Koeffizienten der nicht-diskretisierten Landemissionen und des diskretisierten Niederschlags erkennbar. Werden nur die diskreten Koordinaten verwendet, so ist dem Längengrad ein leichter Jahresgang zuzuordnen. Auch unterscheiden sich die Koeffizienten des Längengrades, im Gegensatz zu denen des Breitengrades, nicht stark voneinander, was ein Hinweis darauf ist, dass eine lineare Modellierung hier genügen könnte. Auffällig ist, dass die Signifikanz der Landemissions- und Niederschlagskoeffizienten von der Jahreszeit abhängen. Sie werden nur im Winter als signifikant von Null verschieden erkannt, wobei die Niederschlagskoeffizienten sich am wenigsten oft von Null unterscheiden. Die Modellgüte schwankt unperiodisch zwischen  $R^2 \approx 10\%$  und  $R^2 \approx 30\%$ . Mit den einfachen linearen Modellen, wo nicht pro Zeitpunkt geschätzt wird, kann eine maximale Modellgüte von circa 20% erreicht werden, wenn die Schiffsemissionen in der größten Gauß'schen Glättung und die Landemissionen und der Niederschlag als diskrete Kovariablen verwendet werden.

Auch die Koeffizienten der additiven Modelle zeigen einen klaren Jahresgang, wobei dieser beim Koeffizienten für den Niederschlag am schwächsten ausgeprägt ist. Zusätzlich stellt sich heraus, dass die Gebiete mit der größten zeitlichen Variation auf der Nordhalbkugel über Asien mit einem Winterminimum und Sommermaximum liegen. Auf der Südhalbkugel hingegen wird fast kein Jahresgang außerhalb des Südpols berechnet. Dort sind auch die zonalen Koeffizienten wesentlich niedriger als über die restlichen Breitengrade hinweg, wo sie annähernd konstant sind. Betrachtet man die räumliche 2D-Verteilung der Koeffizienten der Interaktion von Längen- und Breitengrad, so erkennt man eine gegenläufige Verteilung von Minima und Maxima am Äquator und den mittleren Breiten der Nordhalbkugel. Während über Ozeanien/Indonesien ein Maximum liegt, sind die Koeffizienten über China negativ. Auf der Höhe von Afrika und Europa und über dem Ostpazifik verhält es sich genau gegenteilig. Werden nur die Koordinaten als Kovariablen ins Modell aufgenommen, so zeigen ihre Koeffizienten einen stärkeren Jahresgang als in Modellen mit zusätzlich den Emissions-/Niederschlagsmustern. Der Koeffizient der Schiffsemissionen ist im Sommer am größten, der der Landemissionen hingegen im Winter. Dabei wird der Koeffizient für die Schiffsemissionen häufig durchwegs positiv geschätzt, die Monotonie ist jedoch modellabhängig. Teilweise wird für die höchsten Emissionen wieder ein kleinerer Koeffizient geschätzt, so dass für mittlere Emissionen ein Maximum im Koeffizienten auftritt. Die Koeffizienten der Landemissionen werden teilweise negativ geschätzt. Sie zeigen ein wesentlich breiteres Konfidenzband, welches des öfteren auch die Null umschließt. Dies bedeutet, dass die Schätzung mit einer größeren Unsicherheit behaftet ist und der Einfluss teilweise sehr gering ist. Interessanterweise sind bei vielen der betrachteten Modelle, welche sowohl die Emissions- und Niederschlagsmuster als auch die Koordinaten enthalten, die Konfidenzbänder der Emissions-/Niederschlagskoeffizienten größer als die der Längen- und Breitengrad-Koeffizienten. Mit dem besten Modell lässt sich ein  $R^2$  von circa 40% erreichen, die anderen Modelle liegen bei maximal 35%. Die



Kovariablen und ihre Interaktionen weisen bei einem Test auf Verschiedenheit von Null zum Niveau  $\alpha = 0.05$  durchgängig Signifikanz auf. Die Residuen zeigen das gleiche Verhalten wie bei den Modellen für die Sulfat-Differenz. Die Diagnoseplots, Koeffizientenplots und Signifikanztabellen befinden sich in C.4.7.

In der Gesamtinterpretation zeigt sich auch hier, trotz der geringeren Modellgüte, eine von Null verschiedene Differenz, was zu der Aussage führt, dass die beiden Simulationen sich auch in der Konzentration der Nitrat-Aerosole der untersten Höschicht unterscheiden.

#### 5.2.5.3. Differenz der Ruß-Aerosole (Black Carbon)

Die Koeffizienten der linearen Modelle und der linearen Modelle pro Monat für die Differenz der Ruß-Konzentration verhalten sich ähnlich denen der Nitrat-Differenz. Es ist ein Jahresgang in den Schiffsemissionen und den Breitengraden auf der Nordhalbkugel zu sehen. Zwischen den zeitlichen Verläufen der einzelnen Koeffizienten für die diskrete Version des Längengrades gibt es fast keinen Unterschied. Die Koeffizienten für Niederschlag, Landemissionen und Längengrad schwanken dabei unregelmäßig um die Null, während die Koeffizienten für die Schiffsemissionen und den Breitengrad durchwegs positiv sind. Die Koeffizienten für die Schiffsemissionen sind durchgängig signifikant, wohingegen die für die Landemissionen und den Niederschlag häufiger im Sommer als in den Wintermonaten als signifikant geschätzt werden. Die Modellgüte  $R^2$  schwankt dabei je nach Jahreszeit zwischen 10% und 40%. Im gesamten linearen Modell lässt sich unter Verwendung der Schiffsemissionen mit der kleinsten Gauß'schen Glättung, den per Quantilen diskretisierten Landemissionen und dem durch Rundung diskretisierten Niederschlagsmuster eine maximale Modellgüte von 20% erreichen.

Auch die Koeffizienten der additiven Modelle zeigen den gleichen Unterschied zwischen Nord- und Südhalbkugel in der zeitlichen Variabilität wie bei den Nitrat- und Sulfat-Aerosolen. Während die Koeffizienten des Längen- und Breitengrades für die Südhalbkugel sich kaum mit den Monaten verändern, ist auf der Nordhalbkugel ein klarer Unterschied zwischen Sommer (Maximum) und Winter (Minimum) erkennbar. Die größten Unterschiede sind dabei über Eurasien zu finden, kleinere auch über China. Räumliche Maxima des Koeffizienten der Interaktion von Längen- und Breitengrad liegen über Europa und Ozeanien, räumliche Minima über China und Afrika. Werden nur die Koordinate als Kovariablen verwendet, kann wie bei den Sulfat-Differenzen das räumliche Muster der Koeffizienten der Interaktion von Längen- und Breitengrad, vor allem bei Verwendung von höher aufgelösten Basisfunktionen, mit der räumlichen Verteilung des Schiffsverkehrs in Verbindung gebracht werden. Diese Ähnlichkeit zeigt sich bei Verwendung der Schiffsemissionen als Kovariablen. Der Koeffizient für die Schiffsemissionen ist über alle Modelle hinweg größtenteils positiv, hat im Sommer ein Maxima und kann beinahe als linear angenommen werden. Die Koeffizienten für Landemission und Niederschlag hingegen werden als größtenteils monoton fallend geschätzt und haben wesentlich breitere Konfidenzbänder. Dabei wird meist kein großer Jahresgang geschätzt und wenn doch, so weisen sie im Sommer ein Minimum auf. Beim Signifikanztest für die Koeffizienten des besten Modells zum Niveau  $\alpha = 0.05$  wird der Haupteffekt des Monats als nicht-Signifikant, also als nicht von Null verschieden eingeordnet. Alle anderen Koeffizienten werden als signifikant erkannt. Die Modellgüte liegt über alle (lineare und additive) Modelle mit den Emissions- und Niederschlagsmuster

als Kovariablen hinweg bei durchschnittlich 20%, AIC und BIC geben jedoch den additiven Modellen den Vorzug. Die Diagnoseplots zeigen die gleichen Probleme wie bei den Modellen für die Sulfat- und Nitrat-Konzentrationen auf (Abbildungen C.4.7)

Insgesamt zeigen bei allen drei Konzentrationen die additiven Modelle, welche mehrdimensionale Interaktionen wie Längengrad-Breitengrad-Monate enthalten und räumlich höher aufgelöste Basisfunktionen verwenden, das größte  $R^2$  beziehungsweise das kleinste AIC und BIC aller betrachteten Modelle.

#### 5.2.5.4. Differenz des Wolkenbedeckungsgrades

Die Modellierung der Differenz des Wolkenbedeckungsgrades unterscheidet sich von der der Konzentrationen, da der Wolkenbedeckungsgrad eine wesentlich höhere räumliche Variabilität aufweist.

Im Unterschied zu den Konzentrationen liegen bei der linearen Modellierung des Wolkenbedeckungsgrades die Koeffizienten sämtlicher Kovariablen nahe bei Null und werden beim t-Test nur vereinzelt als signifikant verschieden geschätzt. Auch die Modellgüte liegt durchschnittlich unter 1%. Mit Aufnahme der diskreten Varianten der Emissions- und Niederschlagsvariablen (gleichzeitig) oder der diskreten Koordinaten und ihrer Interaktionen kann ein  $R^2$  von 11-14% erreicht werden. Dabei werden diese Modelle vom AIC und BIC aufgrund ihrer großen Anzahl von Parametern eher abgelehnt.

Betrachtet man die linearen Modelle pro Monat, so ergibt sich ein ähnliches Bild. Die Koeffizienten sämtlicher Kovariablen schwanken um die Null ohne klar erkennbares Muster und die Modellgüte bewegt sich zwischen 1-2%.

Als bestes Modell wurde vom BIC das additive Modell mit Längengrad, Breitengrad und ihrer Interaktion ausgewählt (Abbildung A.19), vom AIC ein wesentlich komplexeres mit den Emissions- und Niederschlagsmustern als Kovariablen, raum-zeitlichen Interaktionen und einer höheren räumlichen Auflösung der Interaktion von Längen- und Breitengrad. Dabei sind AIC und BIC in ihrer Differenz fast gegenläufig. Modelle mit hohem AIC haben ein geringes BIC und umgekehrt. Als Kompromiss können Modelle mit möglichst kleinem AIC und BIC gleichzeitig betrachtet werden, zum Beispiel die Modelle 38, 39, 81, 84 und 96 (Tabelle C.4.1). Diese haben ein durchschnittliches  $R^2$  von 13% und beinhalten Raum-Zeit-Interaktionen, jedoch nur mit geringer räumlicher Auflösung der Basisfunktionen. Die p-Werte der Koeffizienten dieser Modelle sind bis auf die von Längengrad und Monat genauso gering wie die der additiven Modelle für die anderen Variablen. Aber selbst Längengrad und Monat würden bei einem Signifikanztest zum Niveau  $\alpha = 0.05$  als signifikant erkannt werden. Die Residuen zeigen ähnliche Probleme wie die der Modelle der Konzentrationsdifferenzen auf.

Die große Variabilität der Differenz findet man in den sehr breiten Konfidenzbändern wieder, welche häufig die Null beinhalten, außer bei den Koeffizienten für den Breitengrad. Dieser wird über alle Modelle hinweg mit nicht zu breiten Konfidenzbändern geschätzt und ist damit im Bereich des Äquators deutlich negativ und in den mittleren Breiten der Nordhalbkugel noch deutlicher positiv, und damit in beiden Bereichen von Null verschieden. Beim Koeffizienten des Längengrades zeichnet sich ein von Null verschiedenes Minimum für den Bereich über Russland und China ab. Tendenziell ist in der Interaktion von Längen- und Breitengrad ein Minimum über China und ein Maximum über dem Nordpazifik zu erkennen. Eine,

falls vorhandene, zeitliche Variabilität in Form eines Jahresganges tritt nur in den Koeffizienten für den Breitengrad über der Arktis mit einem Minimum im Frühjahr und einem Maximum im Herbst auf. Mit viel Phantasie könnte man dieses Muster mit dem Schiffsemissionsmuster in Verbindung setzen, die Koeffizienten der Schiffs-, Landemissionen und des Niederschlags der Modelle, in welchen die Muster als Kovariablen enthalten sind, weisen jedoch alle sehr große Konfidenzbänder auf, welche die Null enthalten. Es wird von den Modellen also kein Zusammenhang hergestellt.

Insgesamt lässt sich für die Differenz des Wolkenbedeckungsgrades keine so klare Aussage wie für die Differenzen der Konzentrationen treffen. Die durchwegs relativ engen Konfidenzbänder und der kleine p-Wert des Koeffizienten des Breitengrades sprechen für eine relativ sichere Schätzung. Alleine durch die Abweichung des Koeffizienten von Null für die mittleren Breiten der Nordhemisphäre wäre somit ein Modell für die Differenz gefunden und man kann dies als Unterschied zwischen den beiden Simulationen bezeichnen. Die geringe Modellgüte hängt mit der kleinskaligen Variabilität der Differenz zusammen, die mit einem additiven Modell nicht modelliert werden kann. Hier wäre eventuell die Modellierung mit einem Mixed Model (GLMM, GAMM), welches zum Beispiel einen randomisierten Interzept pro Pixel erlaubt oder eine Korrelationsstruktur in den Residuen mit modellieren kann.

### 5.3. Anwendungsgebiet Regressionsmodelle

Regressionsmodelle bieten vor allem die Zusatzinformation über die Stärke eines Zusammenhangs in Form der Koeffizienten des Regressionsmodells an. Wichtig ist dabei die passende Wahl der Kovariablenstruktur und Modellstruktur insgesamt, da Beziehungen verfälscht werden können, wenn nicht alle Eigenschaften der Daten mit modelliert werden. Zum Beispiel wird ein regelmäßig schwankender Zusammenhang wie ein Jahresgang bei einer rein linearen Modellierung nicht berücksichtigt und kann sogar als unzusammenhängend geschätzt werden. Wenn die zu untersuchenden Daten wie in diesem Beispiel räumliche und zeitliche Abhängigkeiten aufweisen, ist die Verwendung von additiven Modellen von Vorteil, vor allem, wenn keine Vorkenntnisse über die Art des Zusammenhangs vorliegen, da die additive Modellierung sich sehr stark an den Daten selber orientiert. Trotzdem ist mit ihr in diesem Fall keine vollständige Modellierung der räumlichen und zeitlichen Korrelation gelungen. Dafür würde sich eine Verbindung mit der Methodik des Kriging anbieten, welche aber einen größeren Computer-Ressourcenbedarf hat. Alternativ kann auch eine Modellierung mit bayesianischen Regressionsmodellen oder gemischten (hierarchischen) Modellen versucht werden. Dies wird ausführlicher in Kapitel 7 beschrieben.

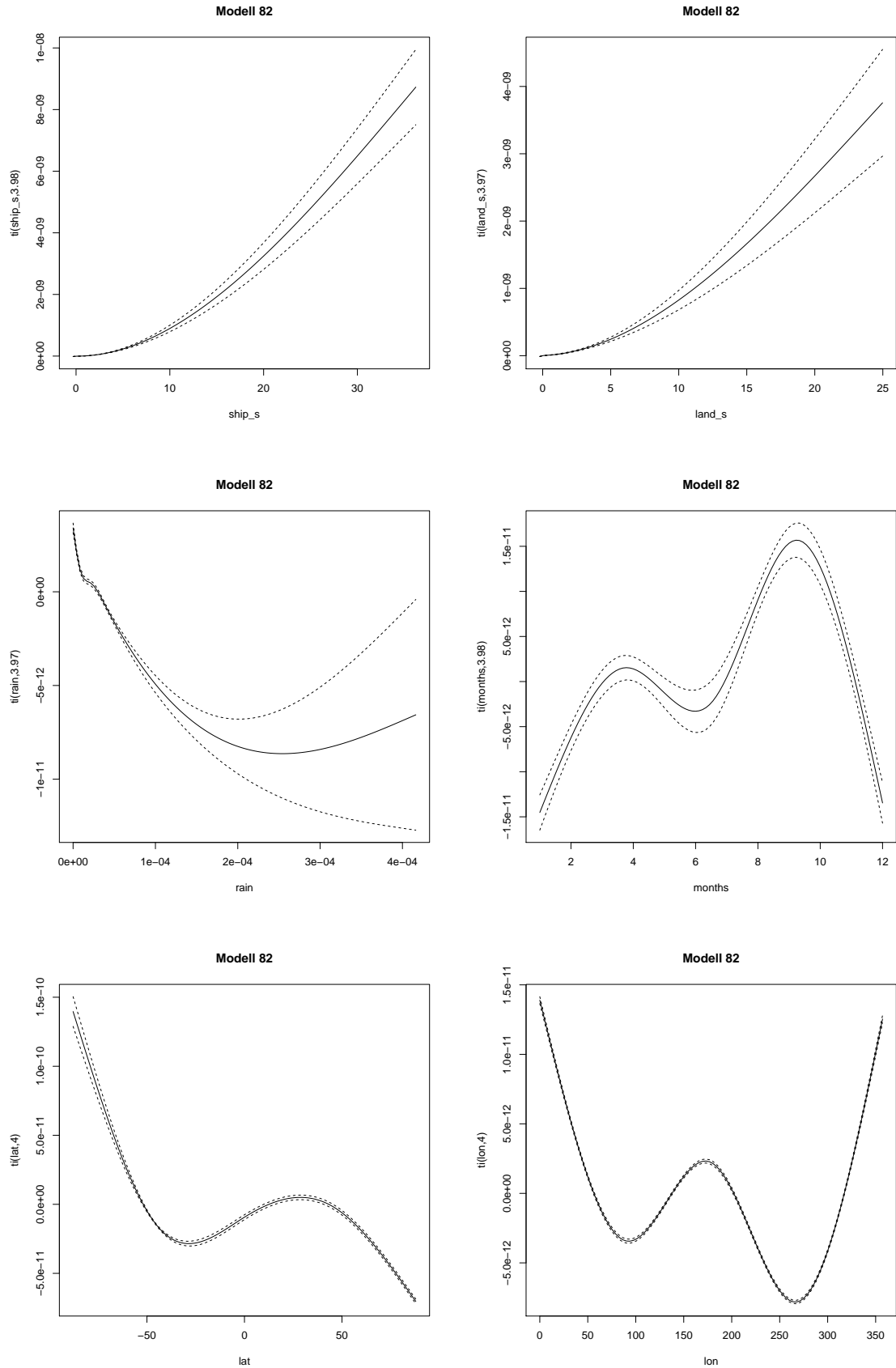


Abb. 5.5.: Darstellung der Haupteffekte des besten additiven Modells für die Differenz der Sulfat-Konzentration in der untersten Höschicht. Dabei ist zu beachten, dass nicht die Koeffizienten an sich gegen die Werte der Variablen, sondern das jeweilige Produkt aus Koeffizient und Variablenwert gegen den Wert der Variablen aufgetragen ist. Die Skala der y-Achse entspricht damit der Skala der zu modellierenden Größe (hier den Konzentrationsdifferenzen)

## 6. Weiterführende Analysemethoden

Die drei in dieser Arbeit angewendeten Methoden sind nicht die einzigen Möglichkeiten, an das Problem der Beschreibung eines Unterschieds heranzugehen. Dieser Abschnitt soll einen kleinen Überblick über weiterführende Verfahren geben und für die in der Auswertung des Beispieldatensatzes nicht angewendeten Definitionen eines Unterschieds verschiedene mögliche Methodiken der Auswertung vorschlagen.

### 6.1. Weitere Möglichkeiten zur Multiplen-Testen-Korrektur

Eine weitere Methode zur Korrektur von Testergebnissen bei multiplem Testen kommt aus der Gehirnforschung, wo sich ebenfalls das Problem des pixelweisen Testens von dreidimensionalen Daten stellt. Diese Methode beruht auf der Annahme eines GMRF (Gaussian Markov Random Field) als Beschreibung für die räumliche Abhängigkeit zwischen den Pixeln. Eine Einführung in die Methoden der fMRI-Analyse ist in [45], eine Einführung in die GMRF-Korrektur findet sich in [74], ein Beispiel für eine Anwendung ist [64].

In [63] wird eine Methode beschrieben, die räumliche Korrelation bei der Multiplen-Testen-Korrektur mit einzubeziehen. Dabei werden die Daten erst per Wavelet-Transformation umgerechnet und dann im Wavelet-Raum die (umformulierten) Hypothesen mit Test-Korrektur getestet. Diese Methodik ist besonders für ein kleines Signal-Rausch-Verhältnis entwickelt worden. Bei ersten Anwendungsversuchen auf den Beispieldatensatz fiel auf, dass hier ähnlich wie bei der PCA das Ergebnis von der Wahl eines Parameters, welcher durch den Anwender zu wählen ist, bestimmt wird.

Es wird schon seit längerem in der Statistik und vor allem in den medizinischen und psychologischen Anwendungsbereichen über die Sinnhaftigkeit der Anwendung einer Korrektur für das multiple Testen diskutiert. Die zwei gegenläufigen Meinungen werden gut durch [60] und [66] wiedergegeben. Einen Überblick über die Problematik liefert [40]. Eng verwandt mit dieser Diskussion ist die Frage, ob überhaupt Hypothesentests und p-Werte verwendet werden sollten. Eine Beschreibung der Situation ist in [23] zu finden.

### 6.2. Bayes und weitere Regressionsmodelle

#### 6.2.1. Weitere Arten von Regressionsmodellen

**Struktur der Kovariablen** Alternativ zur Verwendung von GAMs können MARS-Modelle (Multivariate Adaptive Regression Splines) ausprobiert werden [20]. Diese sind abschnittsweise linear und erlauben somit eine Modellierung von Nichtlinearitäten und vor allem eine flexible Festlegung der Interaktionsstrukturen. Eine weitere Möglichkeit ist die Verwendung von gebrochen-rationalen Funktionen (Fractional Polynomials) zur Darstellung der Nichtlinearität [59]. Als Alternative zur Auswahl

der im Modell enthaltenen Kovariablen per AIC oder BIC gibt es die Möglichkeit, dies den Algorithmus selbst bestimmen zu lassen. In den GAMs geschieht dies durch die Verwendung einer speziellen Basisfunktion (Shrinkage-Basis), für lineare Modelle heißt diese Methode LASSO- und Ridge-Regression. [71]

**Hierarchische Modelle (gemischte Modelle)** Bei der Modellierung der Differenz des Wolkenbedeckungsgrades ist aufgefallen, dass die Differenz räumlich sehr stark variabel ist und die hier betrachteten additiven Modelle diese Variabilität pro Pixel nicht gut wiedergeben können. Es gibt eine Gruppe von Regressionsmodellen, die sogenannten gemischten Modelle (Mixed Models), wo neben den üblichen Kovariablenstrukturen auch die Möglichkeit besteht, Koeffizienten nach Gruppen aufgeteilt zu schätzen, zum Beispiel einen Interzept pro Pixel. Dies würde bewirken, dass die starke kleinskalige Variabilität durch diesen Interzept beschrieben wird. Cressie und Wikle (2011) [16] argumentieren stark für die Modellierung von Raum-Zeit-Daten mit hierarchischen Modellen. Anwendungsbeispiele finden sich in [73, 56, 33, 11, 21], speziell die Anwendung in der Meteorologie wird in [61, 44, 38, 9, 57] beschrieben.

**Modelle für die Varianz, Extremwerte und andere Verteilungsparameter** Für die Fragestellung nach einer Veränderung in der Verteilung zwischen den beiden Simulationen bietet sich die Verwendung von GAMLSS-Modellen (Generalized Additive Models for Location, Scale and Shape) an. Diese können nicht nur eine Vielzahl von Verteilungen modellieren, es kann auch zusätzlich zu einem Regressionsmodell für den Mittelwert eines für die Varianz und die Form aufgestellt werden. Ist man statt an der Beziehung im Mittelwert an dem Einfluss auf andere Quantile, zum Beispiel den Extremwerten (also den höchsten 90% der Werte) interessiert, kann eine Quantil-Regression verwendet werden.

### 6.2.2. Bayesianische Methoden

Der Unterschied zwischen bayesianischen und frequentistischen Methoden ist, dass bei bayesianischen Methoden von Anfang an von einer Verteilung der zu schätzenden Parameter ausgegangen wird. Die Annahme dieser a-Priori-Verteilung ermöglicht es, nach Berechnung des Verfahrens nicht nur einen Punkt-Schätzer für den Mittelwert und die Standardabweichung der interessierenden Parameter zur Verfügung zu haben, sondern eine komplette empirische Verteilung. Aus dieser können dann Aussagen, wie zum Beispiel die Wahrscheinlichkeit, dass der Parameter größer als eine bestimmte Schranke (zum Beispiel Null) ist, direkt berechnet werden. Meist sind die Unsicherheiten in bayesianischen Regressionsmodellen eher konservativ abgeschätzt und bieten somit eine alternative Abschätzung zu den bei komplexeren Modellen nur noch approximativen t-Tests für die Koeffizienten [16]. Ein Beispiel für die Ergebnisse der Anwendung eines linearen bayesianischen Modelles auf die Differenz der Sulfatkonzentration in der untersten Höschicht findet sich in C.4.8.

Bei der Anwendung von semi-Bayesianischen Regressionsmodellen (unter Verwendung des INLA-Algorithmus [39]) zur Beschreibung der räumlichen Korrelation auf den Beispieldatensatz hat sich gezeigt, dass vor allem mit dem Random-Walk-Modell die räumliche Korrelation besser als im additiven Modell beschrieben wer-

den kann (Grafische Auswertung in C.4.8). Allerdings mussten zur erfolgreichen Anwendung vor Berechnung des Modells sämtliche verwendete Variablen, also auch die Konzentrationen und der Wolkenbedeckungsgrad standardisiert werden und der Algorithmus brauchte zu viel RAM-Speicher, um ihn auf den kompletten Datensatz oder auch nur die komplette unterste Schicht, anzuwenden.

### 6.3. Weitere Methoden zur Beschreibung eines Unterschieds

**Forecasting und Image Processing** Gerade im Bereich der Verwendung von Modellen zu Vorhersagezwecken besteht die Notwendigkeit, das Ergebnis des Modells mit tatsächlichen Werten zu vergleichen. Die Güte der Übereinstimmung wird dabei meist durch eine Maßzahl ausgedrückt. Diese Maßzahlen können auch zum Vergleich zweier Modellergebnisse verwendet werden. Je nach Fragestellung gibt es verschiedene Maßzahlen, die sensitiv auf zum Beispiel Extremwerte, räumliche Verschiebung oder Varianzunterschied sind. Cressie und Wikle (2011) [16] in Kapitel 5.7 beschreibt eine Auswahl an Maßzahlen und bietet ausführliche Literaturhinweise. Einen Vergleich verschiedener Methoden bietet [22].

Auch im Gebiet der Bildverarbeitung wurden einige Methoden entwickelt, den Unterschied zwischen zwei Bildern zu quantifizieren. Ein Überblick findet sich in [51, 41], einzelne Methoden werden in [36, 12, 13, 3] beschrieben.

**Machine Learning und Klassifikation** Auch aus dem Bereich der Bildverarbeitung hat sich die Idee entwickelt, über einen Klassifizierungsansatz einen Unterschied zu beschreiben. Dabei wird ein Algorithmus vor die Aufgabe gestellt, zu entscheiden, zu welchem Bild (oder hier: zu welcher Simulation) der Wert des jeweiligen Pixels gehört. Die in Frage kommenden Algorithmen werden unter dem Fach Machine Learning zusammengefasst und beinhalten zum Beispiel neuronale Netze, Support Vector Machines, logistische Regressionsmodelle, Clusteranalysen und weitere. Ein Einführung findet sich in [48].

# 7. Zentrale Aussagen und Methodenüberblick

## 7.1. Vergleich und Beurteilung der Methoden

Die in dieser Arbeit näher betrachteten Methoden werden in Tabelle ?? anhand der in Abschnitt 1.3 definierten Kriterien einander gegenübergestellt. Die Bewertung beruht dabei alleine auf der Anwendung auf den Beispieldatensatz, da keine zusätzlichen Simulationsstudien durchgeführt wurden. Die Aussagen zur jeweiligen Anwendbarkeit ergibt sich dabei aus der Analyse in dem entsprechenden Kapitel. Die Angaben für den Verbrauch der Computerressourcen ist dabei nur approximativ.

### 7.1.1. Ressourcenverbrauch

Verwendet wurden Computer mit 2.4GHz Rechenleistung, und abhängig von der verwendeten Methode, zwischen 8GB und 90GB RAM-Speicher. Dabei hat alleine der komplette Datensatz der Differenzen (mit allen Höhenschichten, Zeitpunkten und Pixeln) eine ungefähre Größe von 8GB, mit Hinzufügen der Kovariablen ist die Größe auf 20GB angewachsen. Die komplexesten der Regressionsmodelle haben trotz der Verwendung der für große Datenmengen optimierten Schätzfunktion *bam* noch einmal zusätzliche 20-30GB beansprucht. Die Rechenzeit war sehr unterschiedlich. Bei den Verfahren, welche auf einer mehrmaligen Anwendung auf Teile des Datensatzes beruhen (zum Beispiel dem Hypothesentesten pro Pixel oder den linearen Modellen pro Monat), verbrauchte die einzelne Anwendung (zum Beispiel auf einen einzelnen Pixel oder einen Monat) relativ wenig Zeit. Da das Verfahren aber sehr oft, zum Beispiel für alle 1,8 Millionen Pixel, angewendet werden musste, addierte sich die kurze Rechenzeit für das einzelne Anwenden auf und führte zu ähnlich langen Zeiten, wie von Verfahren, welche auf den kompletten Datensatz nur einmal angewendet werden.

### 7.1.2. Modellierung von Korrelation in den Daten

In Bezug auf die Anwendbarkeit der untersuchten Methoden in Abhängigkeit der Korrelationsstruktur in den Daten hat sich herausgestellt, dass die einfachen Verfahren, wie der t-Test oder die lineare Regression, nicht in der Lage sind, die Korrelation mit zu berücksichtigen. Allerdings haben auch die komplexeren Verfahren, wie die Kriging-Modelle oder die additiven Regressionsmodelle, keine wesentlich bessere Beschreibung ermöglicht. Eventuelle weitere Arten von Regressionsmodellen wurden in Abschnitt 6.2 kurz vorgestellt.

### 7.1.3. Signifikanz bei niedrigem Signal-zu-Rausch-Verhältnis

Es hat sich bei der hier durchgeführten Analyse herausgestellt, dass im Rahmen der Anwendung auf den Beispieldatensatz die Verfahren des Hypothesentestens pro Pixel und der PCA/EOF-Analyse nicht in der Lage waren, Differenzen im



Wolkenbedeckungsgrad zu detektieren. Die EOF-Analyse hatte auch bei höheren Signal-zu-Rausch-Verhältnissen, wie zum Beispiel bei den Jahresmittel der Differenz der Sulfat-Konzentration, Schwierigkeiten eine signifikante Unterscheidung der Differenz von Null festzustellen. Erst bei Verwendung der ungemittelten oder nicht zentrierten<sup>1</sup> Differenzen oder einer darauffolgenden ICA wurde ein signifikantes räumliches Muster zur Beschreibung der Differenz gefunden. Die Regressionsmodelle hingegen waren bereits auch unter Verwendung einer einfachen Kovariablenstruktur bei der Ermittlung signifikanter Zusammenhänge erfolgreich.

#### **7.1.4. Bewertung**

Im Rahmen der in dieser Arbeit durchgeführten Analyse hat sich gezeigt, dass die EOF-Analyse für die Auswertung eines Differenzunterschieds sowohl mit relativ hohem als auch mit niedrigem Signal-zu-Rausch-Verhältnis nicht geeignet ist. Dies liegt auch an der zeitlichen Variabilität der Daten, welche in einer EOF-Analyse nicht berücksichtigt werden kann. Das Verfahren des t-Tests pro Pixel ist sehr schnell und leicht anzuwenden gewesen. Es müssen jedoch die verschiedenen Voraussetzungen für die Verwendung der jeweiligen Testprozedur vorher überprüft werden und gegebenenfalls über eine Korrektur für multiples Testen nachgedacht werden. In diesem Beispiel hat die Verwendung der Korrektur zu einer besseren und einheitlicheren Interpretierbarkeit der Ergebnisse geführt. Die Regressionsanalyse war bei der Anwendung auf den Beispieldatensatz in der Lage auch ein schwaches Signal (im Wolkenbedeckungsgrad) zu finden und konnte einen Zusammenhang zwischen den Konzentrationsdifferenzen und den Schiffsemissionen herstellen. Dabei war der Aufwand für die Wahl eines guten Modells im Vergleich zum Aufwand für die beiden anderen Methoden wesentlich größer. Insgesamt hängt die Wahl der letztendlich verwendeten Methode davon ab, ob und wie weit die Daten die theoretischen Voraussetzungen erfüllen. Ein weiteres Kriterium ist allerdings auch der zur Verfügung stehenden Zeit für die Analyse.

---

<sup>1</sup>Zentrierung bedeutet das Abziehen des Mittelwertes

Methode	Rechenzeit für kompletten Datensatz <sup>2</sup>	RAM-Speicher	Beschreibung räumliche Korrelation	Beschreibung zeitliche Korrelation	geeignet für nicht-Normalverteilung	geeignet für kleines Signal-Rauschen-Verhältnis
t-Test und Alternativen	15min	< 1GB	nein	nein	ja	nein
multiples-Testen-Korrektur	15min	< 1GB	teilweise <sup>3</sup>	teilweise <sup>4</sup>	teilweise	teilweise
EOF (cdo)	5s	< 1GB	ja	teilweise	nein	nein
EOF (R)	> 30min	2-3GB	ja	teilweise	nein	nein
ICA	30min	2-3GB	ja	teilweise	ja	nein
lineare Regression	10min	1-2GB	teilweise	teilweise	mit Erweiterungen	ja
Geostatistik	2-3h <sup>5</sup>	2-3GB	ja	ja	mit Erweiterung	ja
additive Regression	10min bis >5h	1-30GB	größtenteils	größtenteils	mit Erweiterung	ja
Bayesianische Regression	10min bis >5h	1-90GB <sup>6</sup>	ja	ja	mit Erweiterung	ja

Tabelle 7.1.: Vergleich der verwendeten Methoden anhand der anfangs aufgezählten Kriterien

---

<sup>2</sup>gemessen auf einem Computer mit 1.6GHz

<sup>3</sup>Eine kurze Beschreibung von zwei Methoden, welche die räumliche Abhängigkeit zwischen p-Werten berücksichtigen, findet sich in Abschnitt 6.1

<sup>4</sup>nicht bei Mittelung über die Zeit

<sup>5</sup>Bei der Verwendung geostatistischer Modelle muss bedacht werden, dass vor Durchführung des Krigings, also der eigentlichen Berechnung, die gewählte Kovarianzfunktion (Erklärung in 5.1.3) an das empirische Variogramm (Definition in 2.2.2.3) angepasst werden sollte. Dabei ist vor allem die Berechnung eines raum-zeitlichen Variogramms sehr zeitaufwendig. Für die hier ausgewerteten Variablen dauerte die Berechnung pro Variable circa 5-6 Stunden.

<sup>6</sup>Der hohe RAM-Speicherverbrauch geht jedoch eventuell auf eine Fehlspezifikation im verwendeten Algorithmus zurück.

## 7.2. Leitfaden für die Analyse

Abschließend soll durch eine strukturierte Auflistung der Ergebnisse eine Entscheidungshilfe zur Methodenfindung gegeben werden. Dieser Leitfaden ist aus der Erfahrung bei der Auswertung des Beispieldatensatzes entstanden und erhebt damit keinen Allgemeingültigkeitsanspruch.

Der erste Schritt besteht in der Festlegung der Fragestellung, aus welcher sich dann eine Auswahl an Analysemethoden ergibt. Wie in Kapitel 1 beschrieben, muss bei der allgemeinen Fragestellung nach einem Unterschied erst definiert werden, wodurch sich der Unterschied physikalisch manifestiert, zum Beispiel durch eine Differenz oder eine räumliche Verschiebung von Mustern. Davon ausgehend können dann verschiedene Analysemethoden in Betracht gezogen werden.

1. Differenz: t-Test (eventuell) mit multipler Testen-Korrektur, Forecasting Scores (Erklärung in 6.3), Regressionsmodelle, Klassifikation (Erklärung in 6.3)
2. Veränderung in der Verteilung (andere Varianz oder Schiefe oder Extremwerte): GAMLSS- und Quantil-Regression (6.2)
3. Räumliche Verschiebung, Ausdehnung, Drehung oder ähnliches: Forecasting scores, Image Registration (Erklärung in 6.3), Cluster-Analysen
4. Zeitliche Verschiebung von räumlichen Mustern: POP-Analyse, EEOF-Analyse, Hilbert-EOF-Analyse (Erklärung in [? 69])
5. Unterschiedlicher zeitlicher Verlauf: Zeitreihentheorie, Regression mit Lag (Erklärung in 5.1.4)
6. Unterschied in mehreren Variablen: Multivariate Regression, Fingerprinting-Methode ([25, 26])

Im zweiten Schritt werden die Daten auf ihre Eigenschaften wie Wertebereich, Verteilung, räumliche und/oder zeitliche Korrelation untersucht, damit diese bei der Auswahl des passenden Analyseverfahrens mit berücksichtigt werden können. Welche Maße beziehungsweise Diagnosegrafiken dafür geeignet sind, ist in Kapitel 2 beschrieben. Außerdem können für sehr großen Datensätzen hier bereits Möglichkeiten für und Voraussetzungen von Verfahren zur Dimensionsreduktion überprüft werden. Ist ein Analyseverfahren gefunden worden, so kann es angewendet und gegebenenfalls in iterativen Schritten noch verfeinert werden. Drei Beispiele für eine solche Anwendung liefern die Kapitel 3, 4 und 5. Kommen mehrere Analyseverfahren in Frage, so können deren Ergebnisse nach ihrer Anwendung miteinander verglichen werden und gemeinsam in der Gesamtinterpretation verwendet werden. Dadurch ergibt sich mitunter eine detailliertere Auswertung.

## 7.3. Zusammenfassende Auswertung des Beispieldatensatzes

Es wurden die Differenzen von vier physikalischen Größen (Konzentration von Sulfat-Aerosolen, Nitrat-Aerosolen, Ruß und dem Wolkenbedeckungsgrad) zwi-

schen zwei Klimamodellsimulationen betrachtet. Die beiden Simulationen unterscheiden sich durch die zusätzliche Aufnahme der Emissionen des Schiffsverkehrs in ein Modell. Es wurden zur Auswertung die monatlichen Mittelwerte von 10 Jahren verwendet. Dabei zeigten sowohl die Daten an sich, als auch die Differenzen zwischen den Simulationen, eine starke räumliche und zeitliche Korrelation. Insgesamt war die Differenz in den untersten Schichten am größten, weshalb sich die weitere Auswertung hauptsächlich auf diese konzentriert. Bereits die Auswertung von t-Tests mit multipler-Test-Korrektur zum Niveau  $\alpha = 0.05$  pro räumlicher Koordinate für die jeweiligen Jahresmittelwerte zeigte einen großen Anteil an signifikanten Zurückweisungen der Nullhypothese für die Konzentrationen, was ein Hinweis auf eine von Null verschiedene Differenz ist. Dabei zeigten auch die Testergebnisse eine räumliche Korrelation. Die Testergebnisse der Differenz des Wolkenbedeckungsgrades zeigten für keinen Pixel eine Signifikanz an. Durch die Verwendung von additiven Regressionsmodellen konnte die Differenz noch präziser beschrieben werden. Es zeigte sich vor allem für die Differenz der Sulfat-Konzentration eine starke Abhängigkeit vom Jahresverlauf mit Maxima im Sommer und Minima im Winter. Räumlich gesehen verteilen sich die Maxima der Differenz auf die Nordhalbkugel über den Ozeanen und können somit mit den Schiffsemissionen in Verbindung gebracht werden. Bei einer räumlich höher aufgelösten Modellierung konnten teilweise auch Zusammenhänge mit den Emissionen über China und Europa gefunden werden. Die anderen beiden Konzentrationen verhalten sich ähnlich, wobei sich der Einfluss der Schiffsemissionen als nicht ganz so stark herausgestellt hat. Für die Differenz des Wolkenbedeckungsgrades sind die hier betrachteten Modelle nicht besonders gut, zeigen aber eine schwache Abhängigkeit vom Breitengrad auf. Insgesamt lässt sich also sagen, dass sich die beiden Simulationen in der Differenz definitiv unterscheiden, wobei bei den Aerosol-Konzentrationen die Differenz in direkten Zusammenhang mit den Schiffsemissionen gebracht werden kann.

# A. Weitere Abbildungen

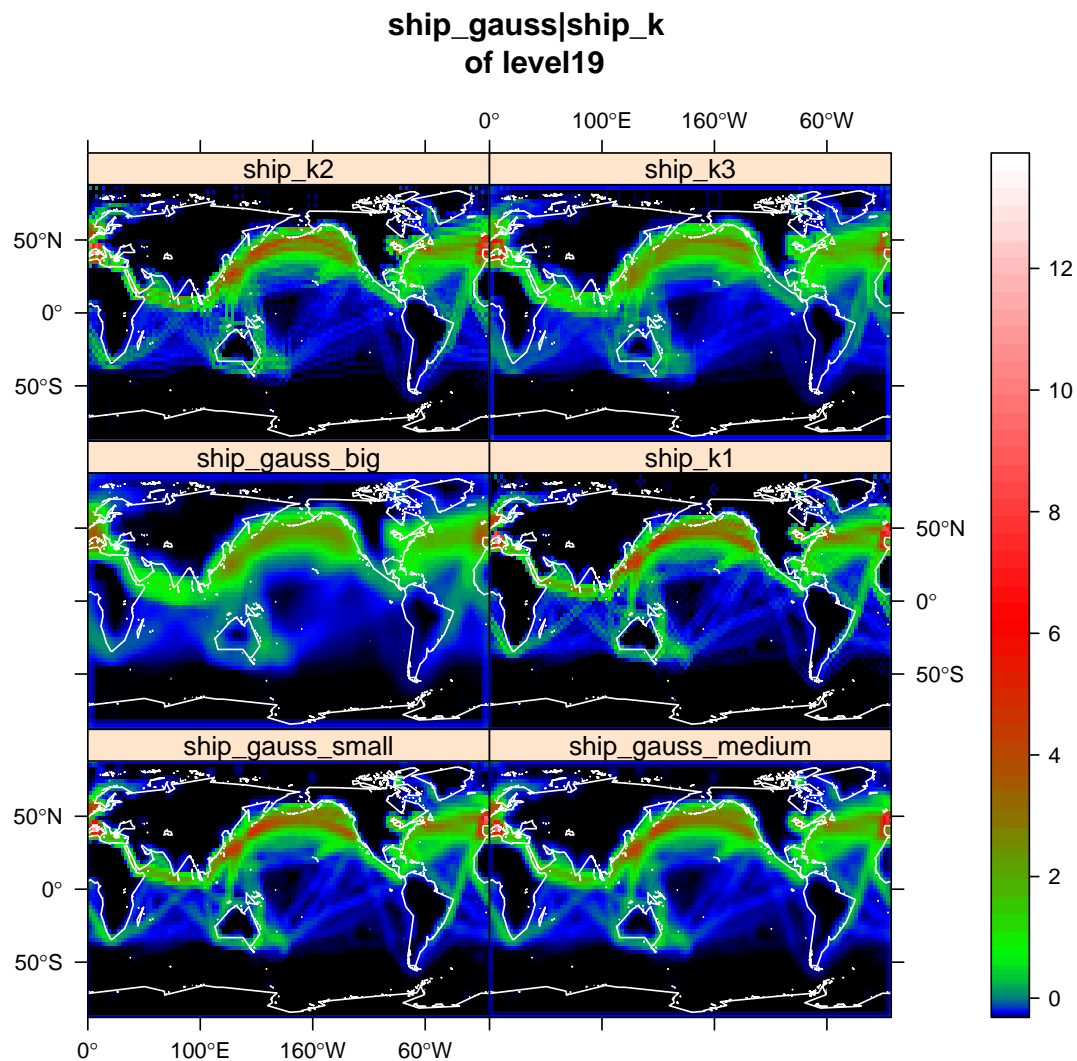


Abb. A.1.: Räumliche Verteilung der geglätteten standardisierten Schiffsemissionen für Juli

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.5e-11	8.3e-13	42	0

	edf	Ref.df	F	p-value
ti(ship_s)	4	4	4.1e+02	0
ti(land_s)	4	4	2.2e+02	3e-188
ti(rain)	4	4	2.5e+02	1.1e-217
ti(months)	4	4	1.4e+02	7.9e-117
ti(lat)	4	4	1.3e+04	0

Continued on next page

	edf	Ref.df	F	p-value
ti(lon)	4	4	6.4e+03	0
ti(ship_s,months)	15	16	81	2.5e-259
ti(land_s,months)	12	13	57	2.1e-151
ti(rain,months)	13	13	34	3.9e-87
ti(lat,months)	16	16	2.5e+03	0
ti(lon,months)	16	16	5.5e+02	0
ti(lat,lon)	16	16	1.6e+03	0
ti(lat,lon,months)	64	64	2.5e+02	0
ti(ship_s,lat)	16	16	3.1e+03	0
ti(ship_s,lon)	16	16	7.2e+02	0
ti(land_s,lat)	16	16	1.3e+03	0
ti(land_s,lon)	16	16	6.8e+02	0
ti(rain,lat)	13	13	1.1e+02	8.6e-289
ti(rain,lon)	16	16	5.6e+02	0
ti(ship_s,lat,months)	63	64	1.1e+03	0
ti(ship_s,lon,months)	63	64	3.9e+02	0
ti(land_s,lat,months)	57	59	1.9e+02	0
ti(land_s,lon,months)	53	55	91	0
ti(rain,lat,months)	61	63	49	0
ti(rain,lon,months)	63	64	1.6e+02	0

Tabelle A.2.: Signifikanztabelle des besten additiven Modells für die Sulfat-Konzentration der untersten Höschicht. In Spalten 2 und 3 sind für die geglätteten Terme die geschätzte Anzahl an Freiheitsgraden, die bei der Schätzung des jeweiligen Terms verwendet wurden, eingetragen; in Spalte 4 der Wert der Teststatistik für die Durchführung eines approximativen Tests auf Verschiedenheit von Null. In Spalte 5 ist der zum t-Test zugehörige p-Wert aufgetragen, welcher mit dem gewählten Signifikanzniveau verglichen wird. Für den Fall geglätteter Kovariablen (gekennzeichnet durch den Zusatz ti()) ist dieser p-Wert nur approximativ gültig. Besonders wenig aussagekräftig ist er, wenn er in der Größenordnung des Signifikanzniveaus liegt.

**p-Werte für Shapiro-Wilk-Test  
der Jahresmittel,  
Signifikanzniveau 0.05, Level 19**

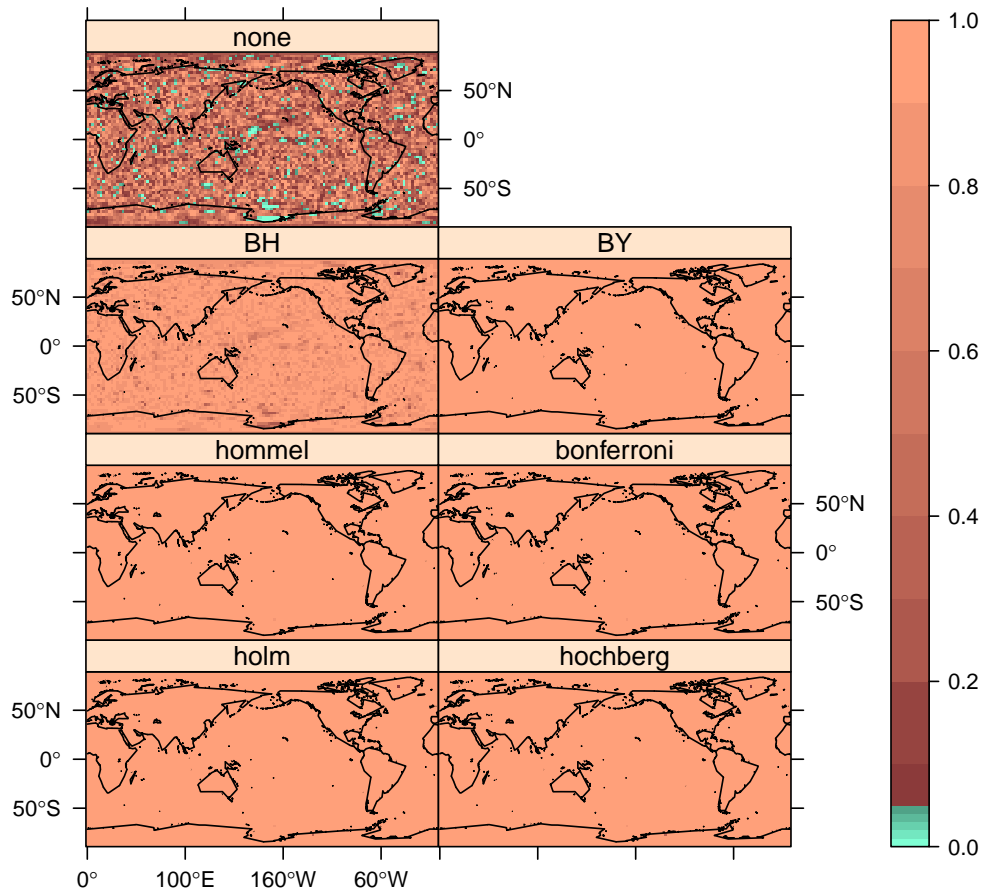


Abb. A.2.: Räumliche Verteilung der p-Werte für einen Shapiro-Wilk-Test auf den Jahresmittelwerten der Differenz der Sulfat-Konzentration in der untersten Höschicht. Wird ein Signifikanzniveau von 5% angenommen, so zeigt die türkise Farbe eine Ablehnung der Nullhypothese zu diesem Niveau an, was hier eine Ablehnung der Normalverteilungsannahme bedeutet.

## Shapiro–Wilk–Tests für

Level 19,  
alle Zeitpunkte

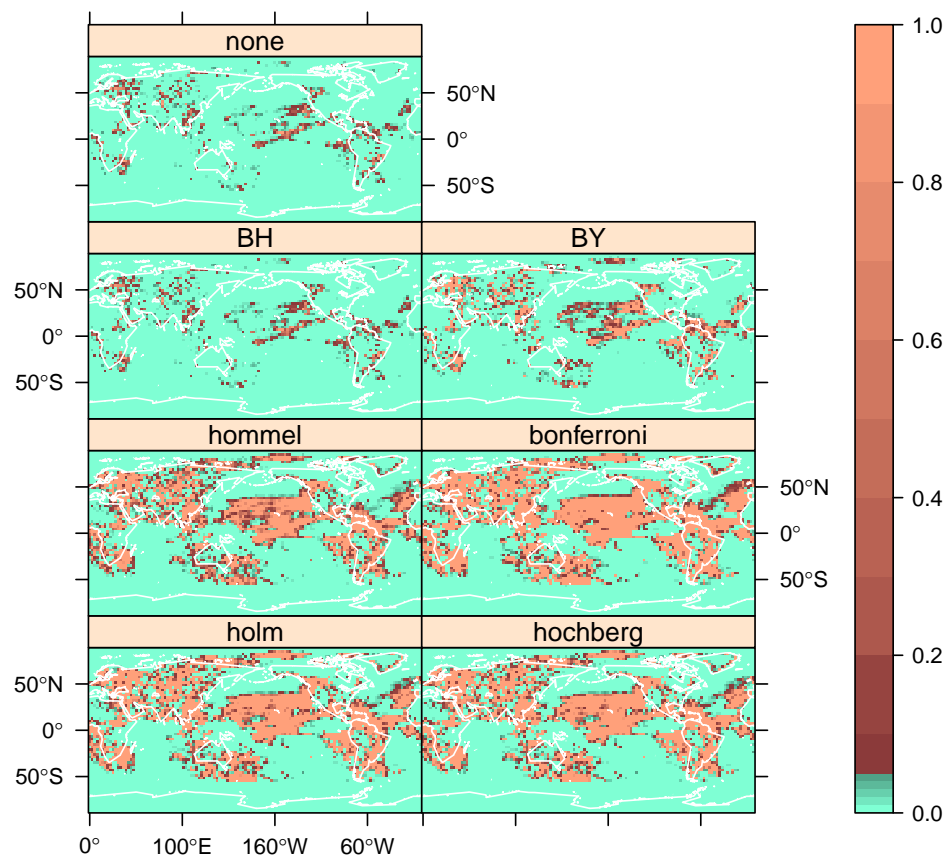


Abb. A.3.: Räumliche Verteilung der p-Werte für einen Shapiro-Wilk-Test der Differenz der Sulfat-Konzentration in der untersten Höschicht. Wird ein Signifikanzniveau von 5% angenommen, so zeigt die türkise Farbe eine Ablehnung der Nullhypothese zu diesem Niveau an, was hier eine Ablehnung der Normalverteilungsannahme bedeutet.



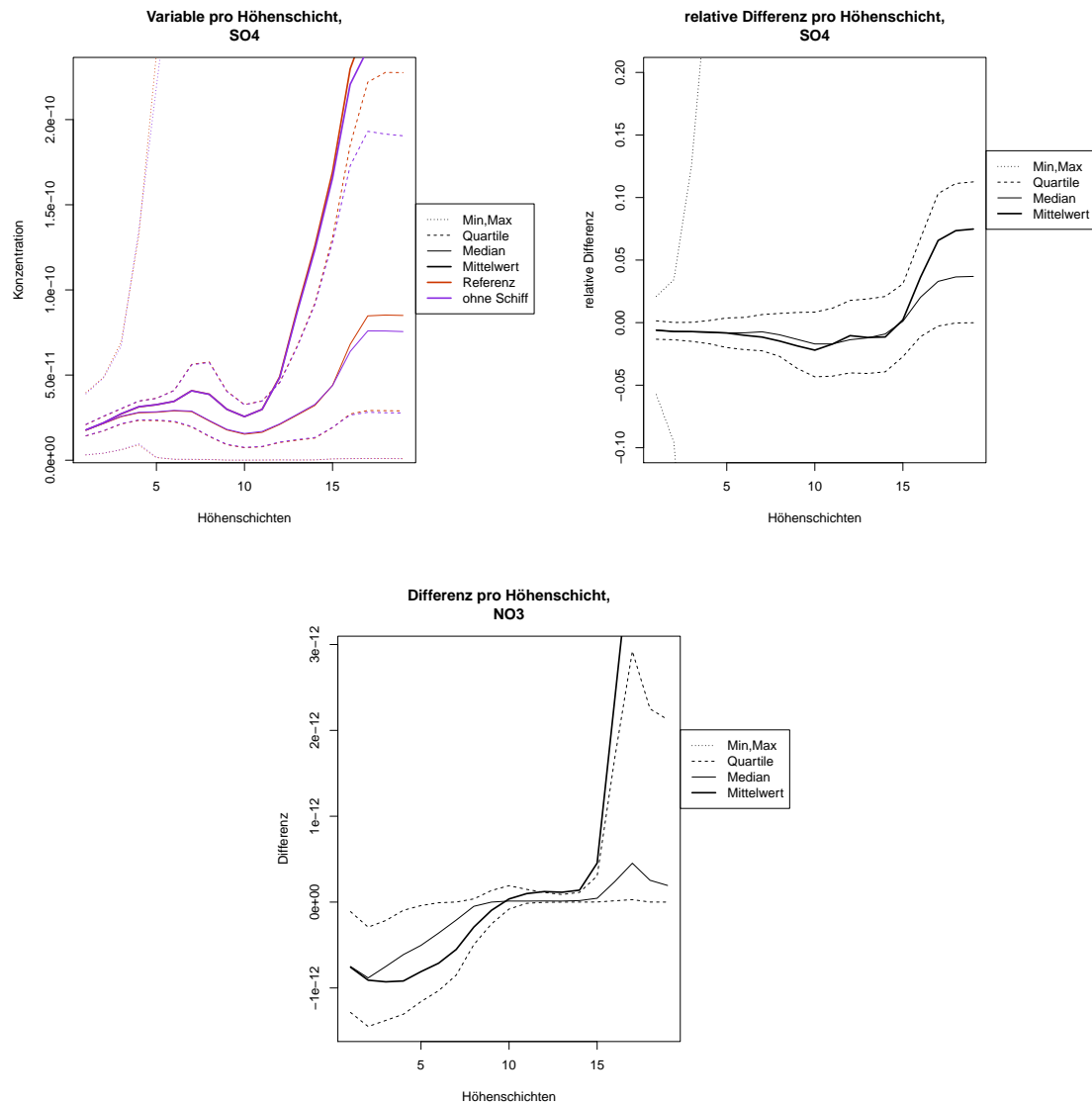


Abb. A.4.: Variation der 5-Punkte-Zusammenfassung (Minimum, erstes Quartil, Median, Mittelwert, drittes Quartil, Maximum) des Sulfat-Aerosols in [mol/mol] mit den Höhenschichten

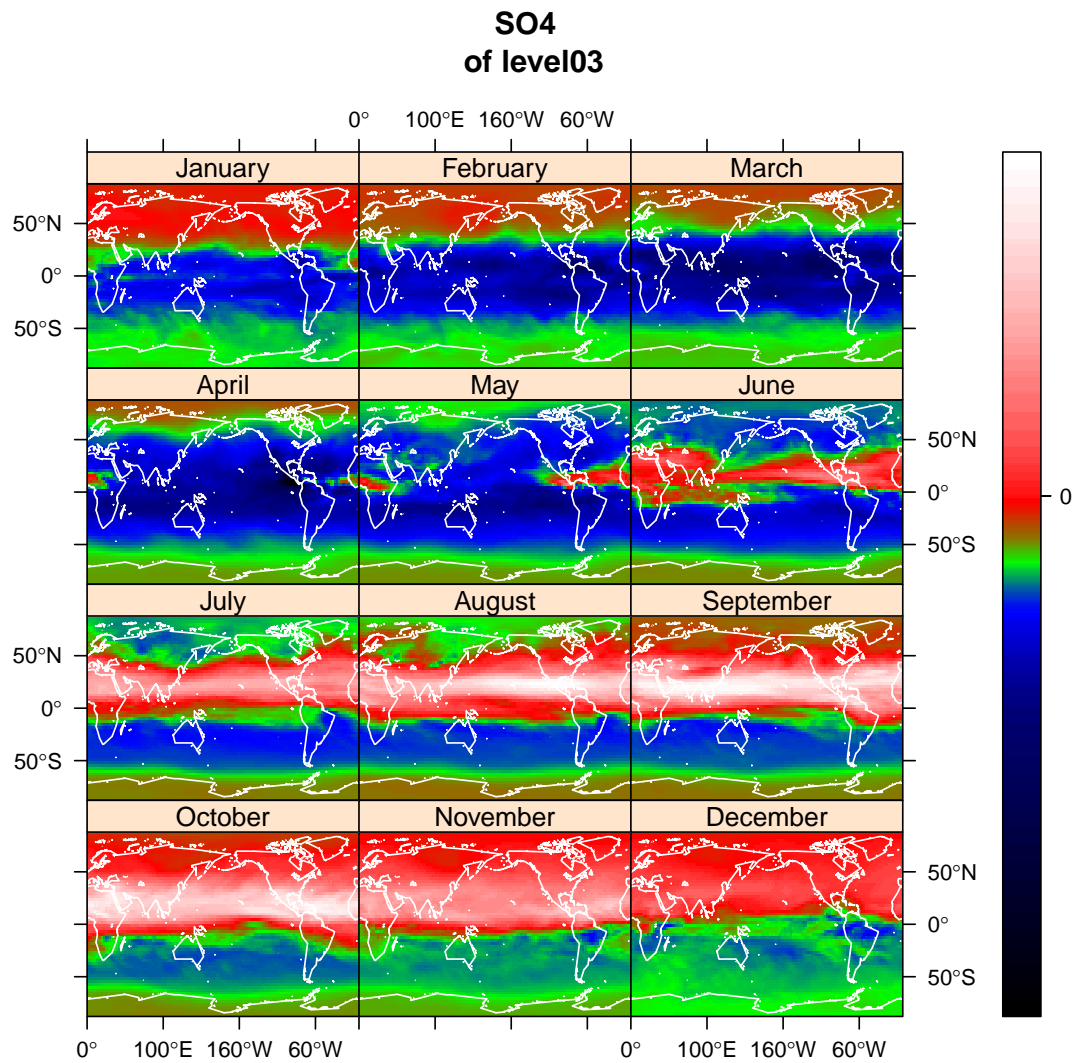


Abb. A.5.: Räumliche Verteilung der Differenz der Sulfat-Konzentration in  $[\text{mol/mol}]$ , gemittelt über die jeweiligen Monate in der oberen Modellatmosphäre (entspricht ungefähr der Stratosphäre)

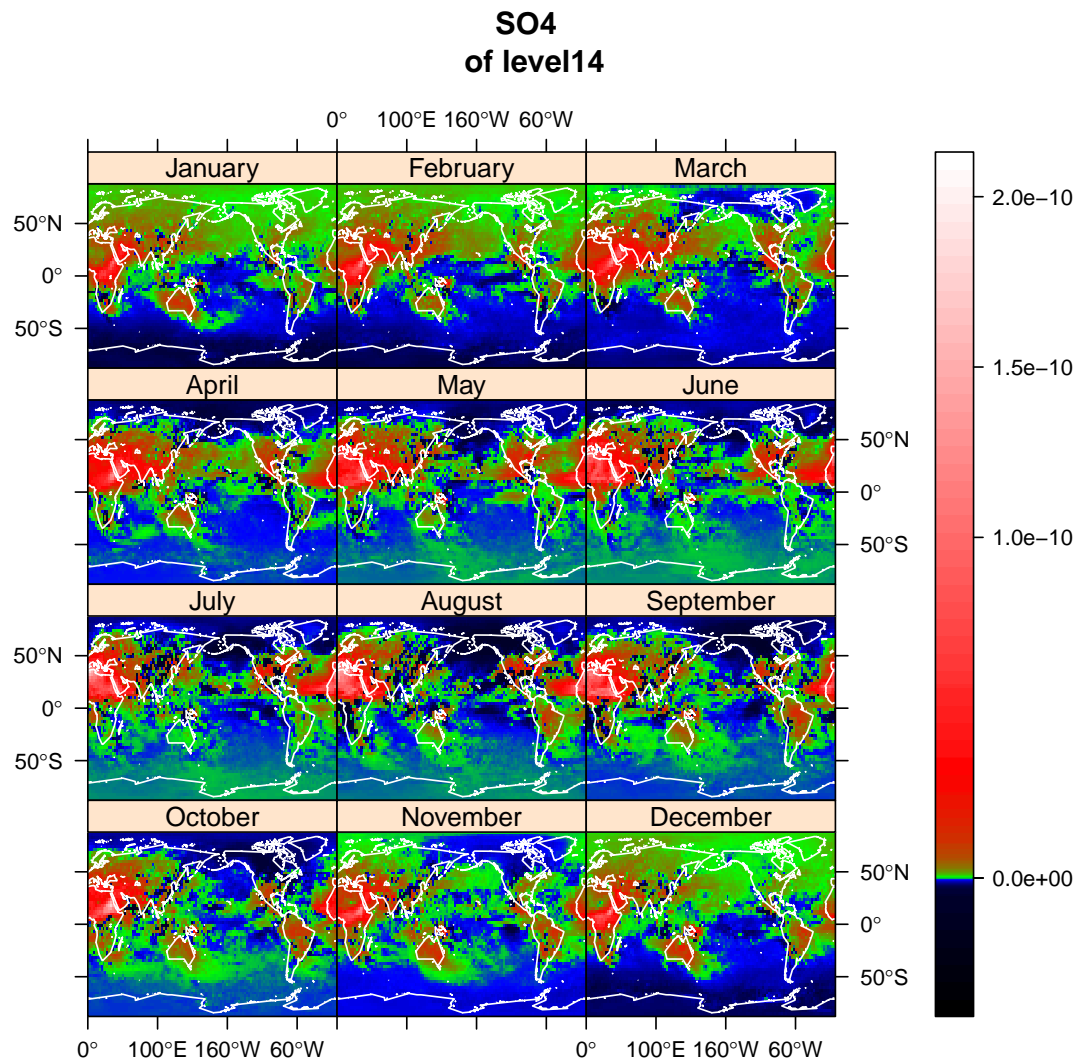


Abb. A.6.: Räumliche Verteilung der Differenz der Sulfat-Konzentration in [mol/mol], gemittelt über die jeweiligen Monate in der mittleren Modellatmosphäre (entspricht ungefähr der freien Troposphäre)

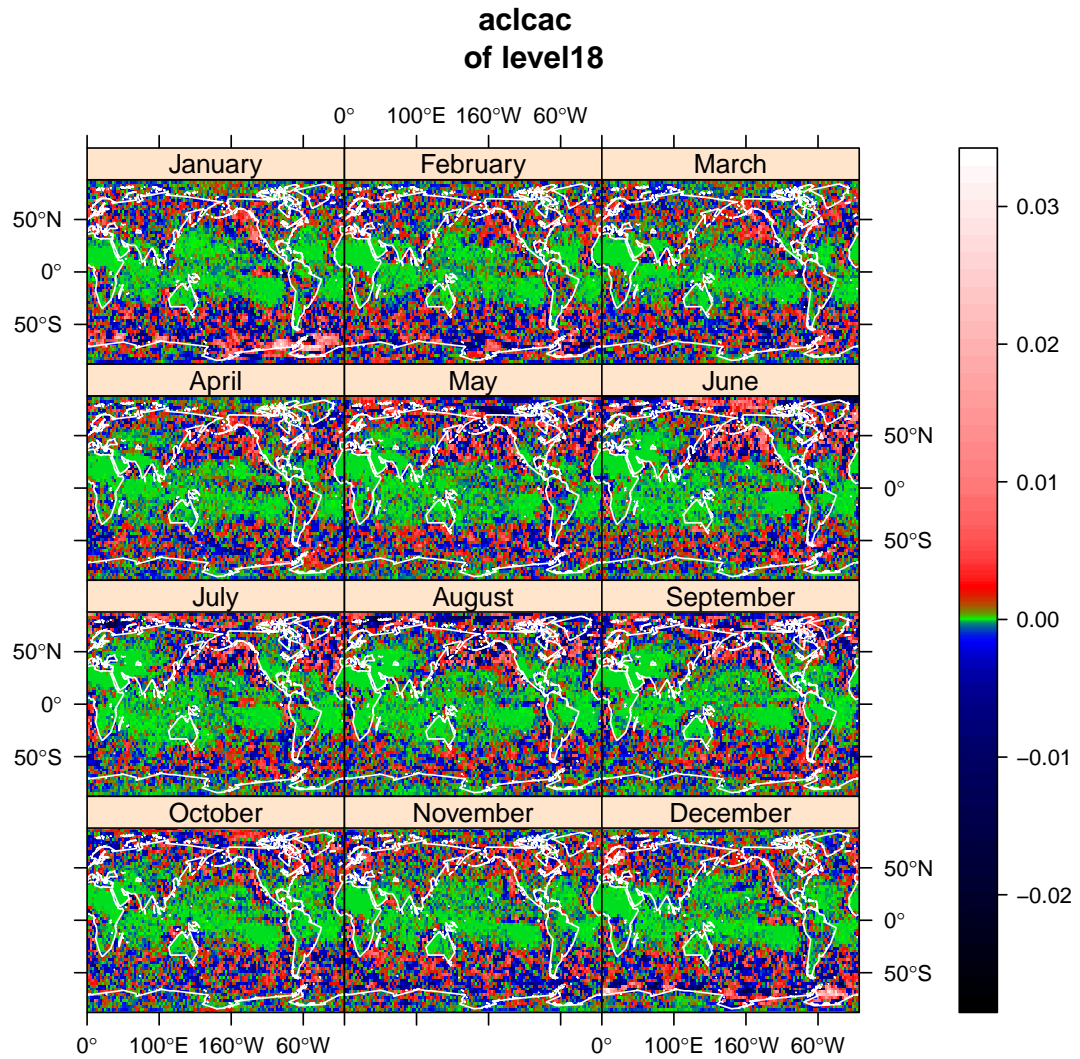


Abb. A.7.: Räumliche Verteilung der Differenz des Wolkenbedeckungsgrades, gemittelt über die jeweiligen Monate in der unteren Modellatmosphäre (entspricht ungefähr der planetaren Grenzschicht)

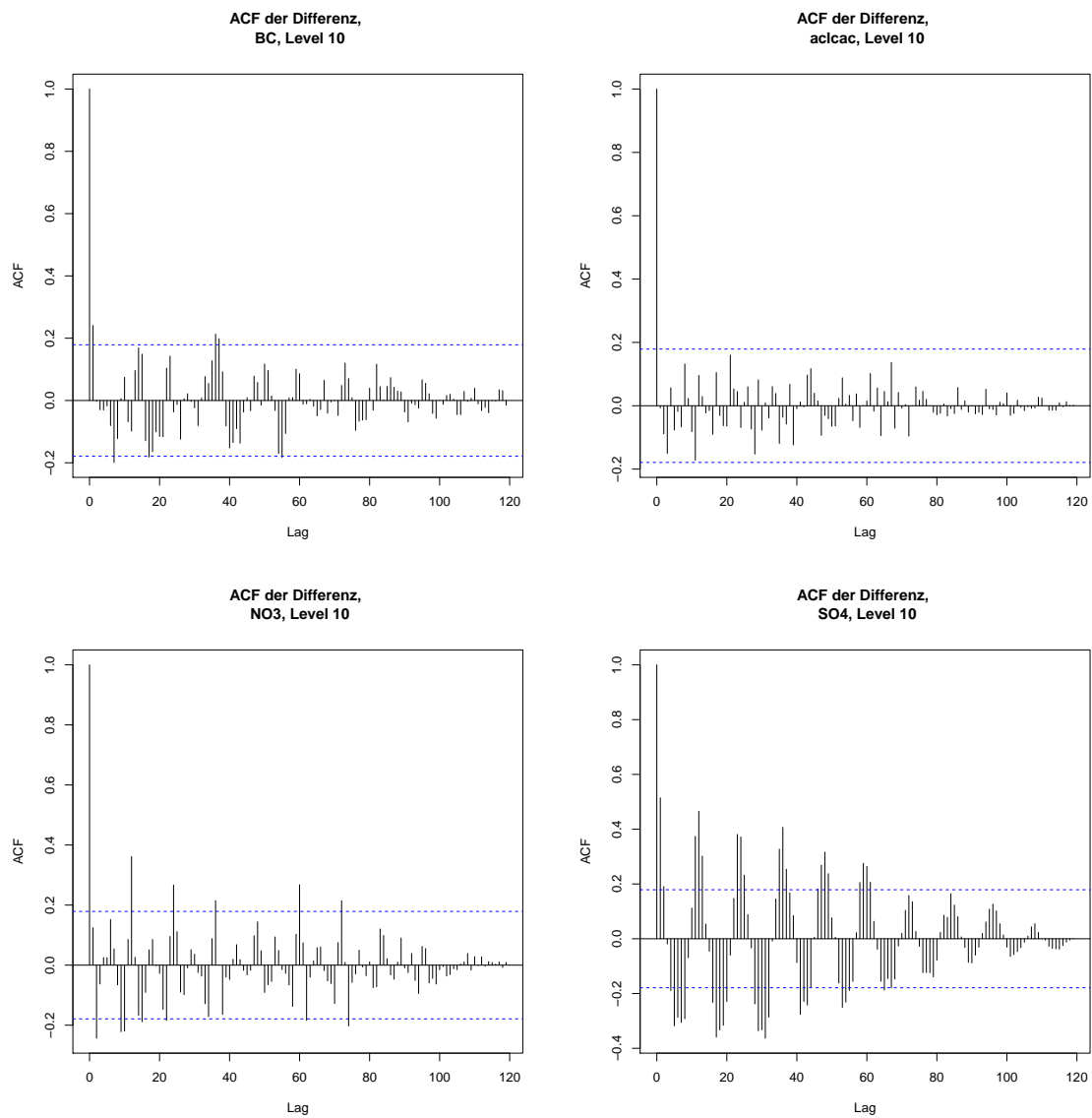


Abb. A.8.: empirische Autokorrelationsfunktion der globalen Mittelwerte der Differenz der vier Variablen für die unterste Höhengschicht

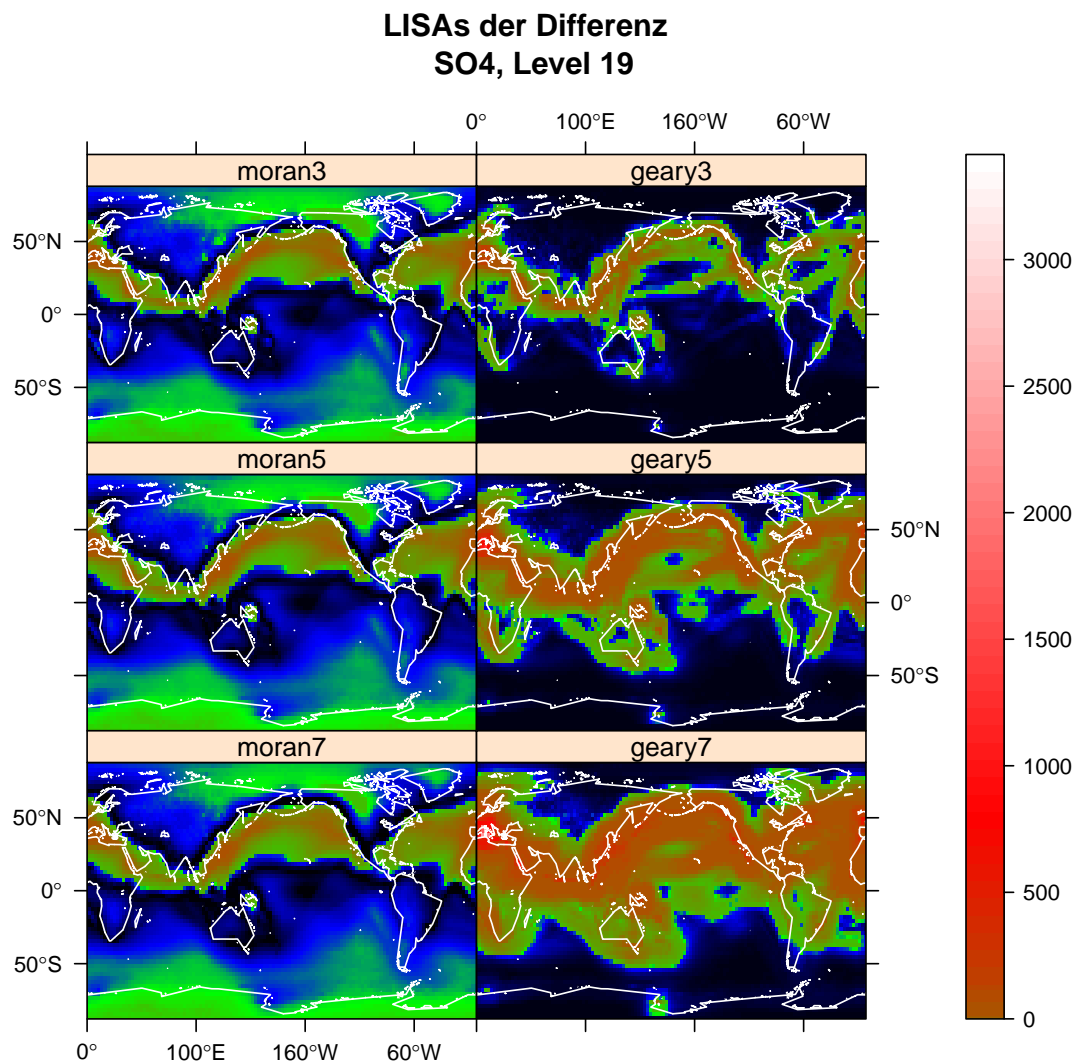
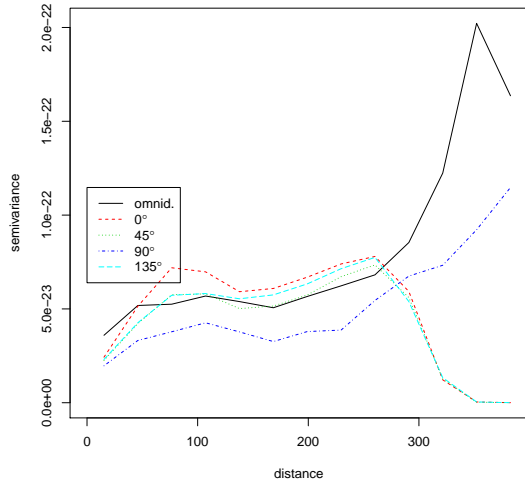
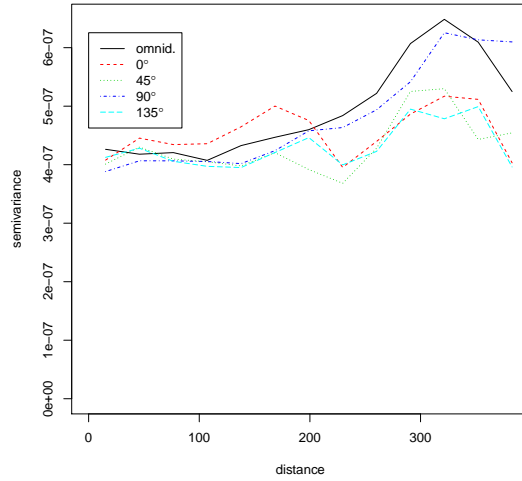


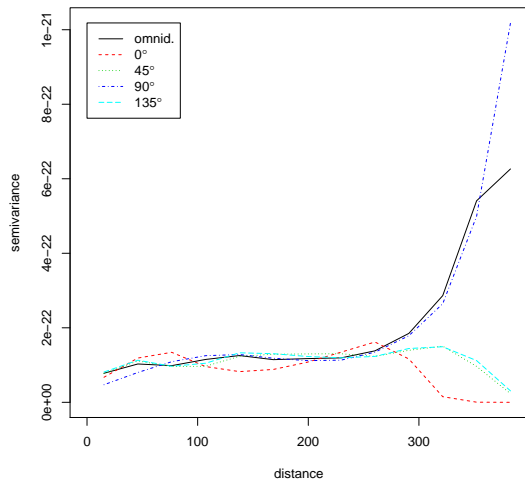
Abb. A.9.: Lokale Indikatoren für räumliche Abhängigkeiten (LISAs) der zeitlichen Mittelwerte der Differenz der Sulfat-Konzentration für die unterste Höschicht mit variierender Größe des lokalen Einflussgebietes (3 Pixel, 5 Pixel und 7 Pixel). Je extremer der Wert, desto größer der Einfluss der Region auf die Globalstatistik.



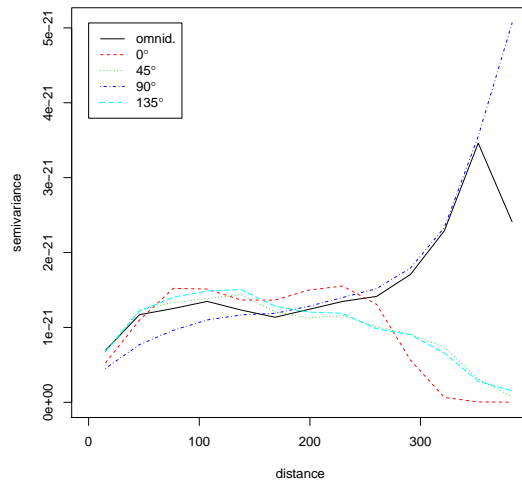
(a) Variogram BC(Ruß)



(b) Variogram cloudcover



(c) Variogram NO3



(d) Variogram SO4

Abb. A.10.: empirisches räumliches Variogramm der globalen Mittelwerte der Differenz der vier Variablen für die unterste Hörschicht in verschiedenen Richtungen ( $0^\circ$  = meridional (Nord-Süd),  $90^\circ$  = zonal (Ost-West),  $45^\circ$  = NordOst-Südwest,  $135^\circ$  = NordWest-SüdOst) und mit allen Richtungen (omnid. = omnidirektional). Je höher die Semivarianz, desto niedriger die räumliche Assoziation. Die räumliche Distanz ist dabei in Grad angegeben. Die Punktepaafe wurden mit einer Toleranzgrenze von  $\pm 10^\circ$  ausgewählt.

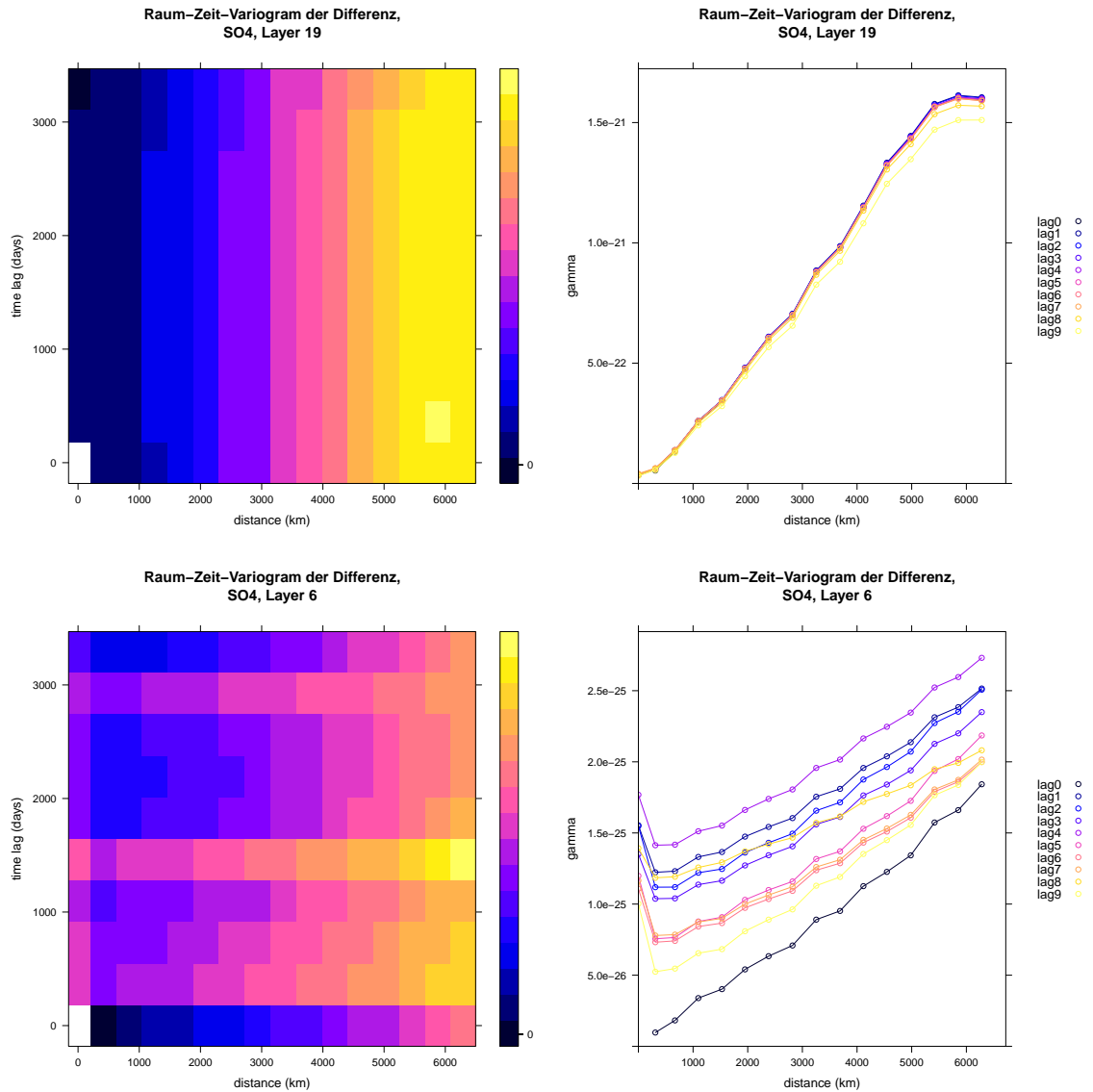


Abb. A.11.: empirisches raum-zeitliches Variogramm der Sulfat-Differenz für die unterste und dritt-oberste Höhengschicht in zwei verschiedenen Darstellungen. Je höher die Semivarianz  $\gamma$ , desto kleiner die raum-zeitliche Assoziation.



**p-Werte für Wilcox-Test der Jahresmittel  
Signifikanzniveau 0.05, Level 19**

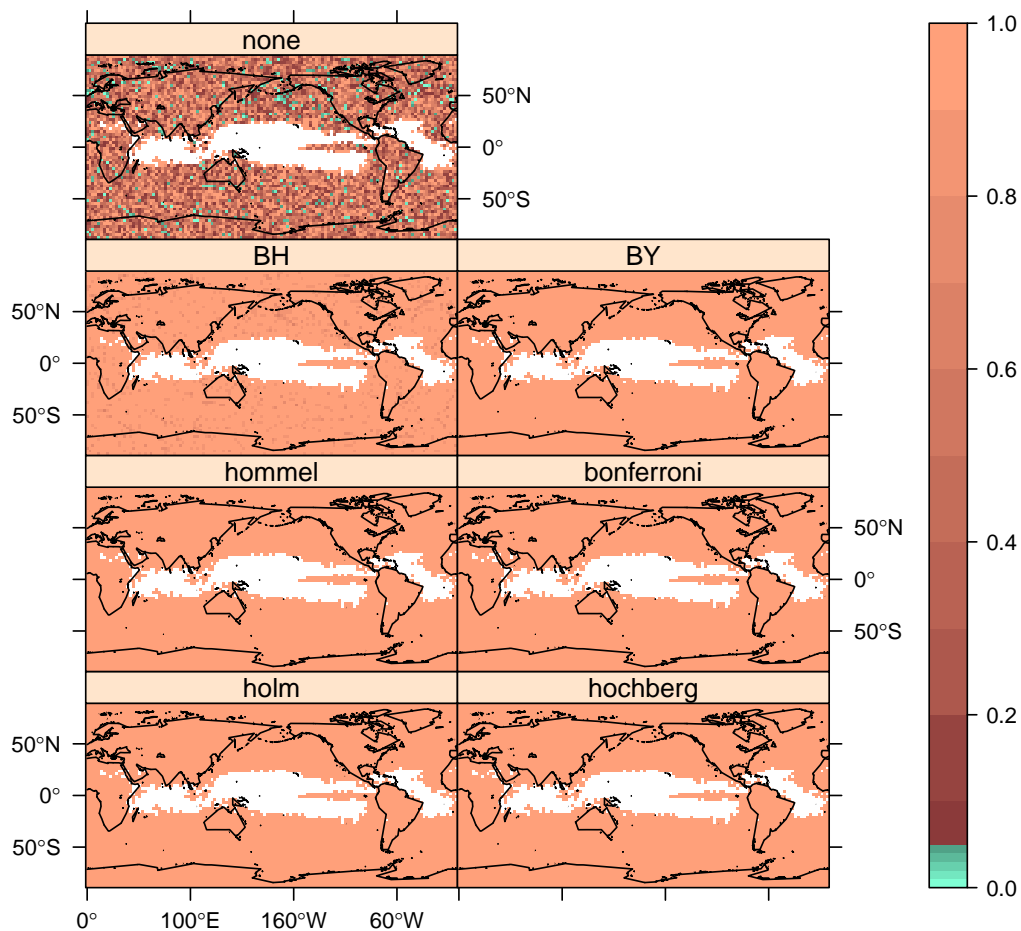


Abb. A.12.: Räumliche Verteilung der p-Werte für einen zweiseitigen Wilcox-Vorzeichen-Rang-Test auf den Jahresmittelwerten der Differenz des Wolkenbedeckungsgrades in der untersten Höhengschicht. Wird ein Signifikanzniveau von 5% angenommen, so zeigt die türkise Farbe eine Ablehnung der Nullhypothese zu diesem Niveau an, was hier eine Ablehnung der Normalverteilungsannahme bedeutet.

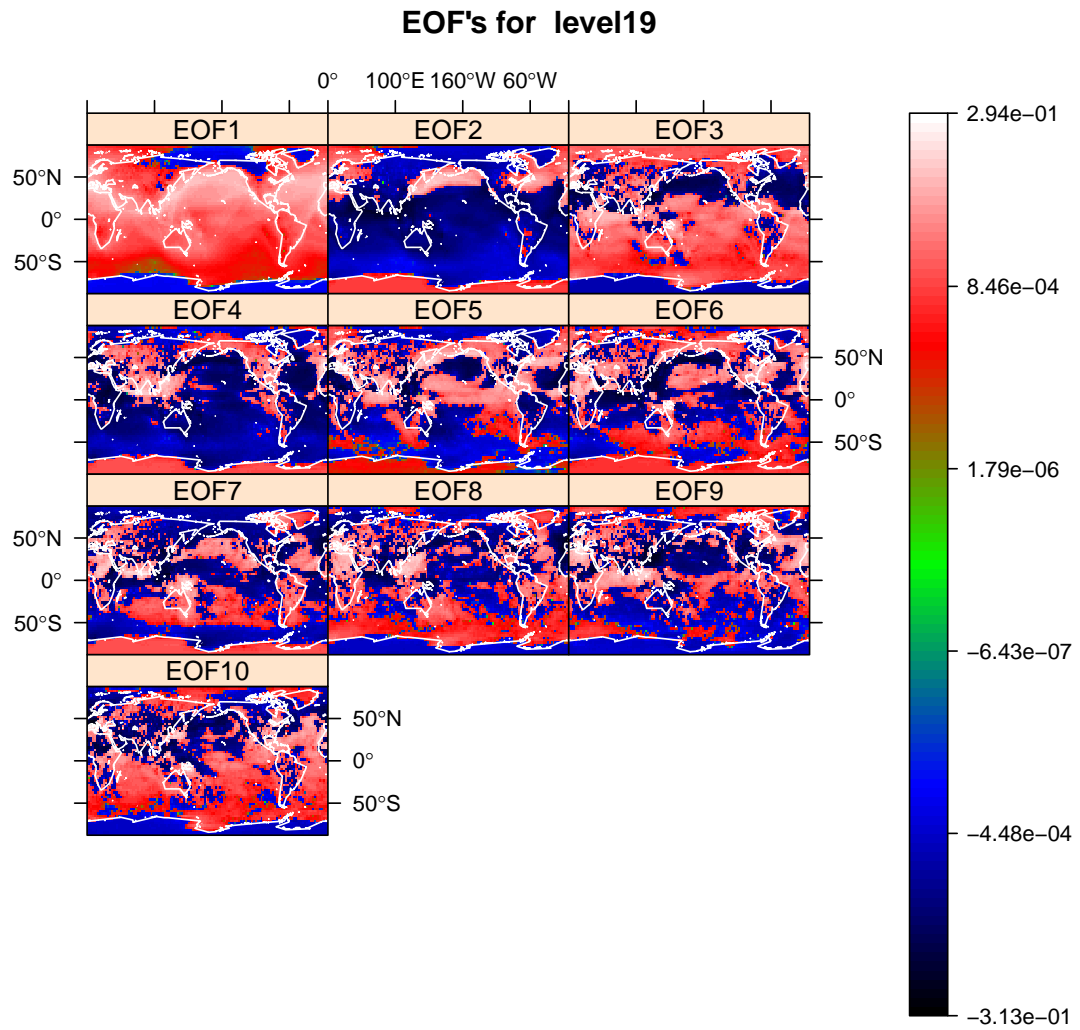


Abb. A.13.: Räumliche Eigenvektoren der Sulfat-Differenz der untersten Höhen-  
schicht, ohne Zentrierung oder Standardisierung

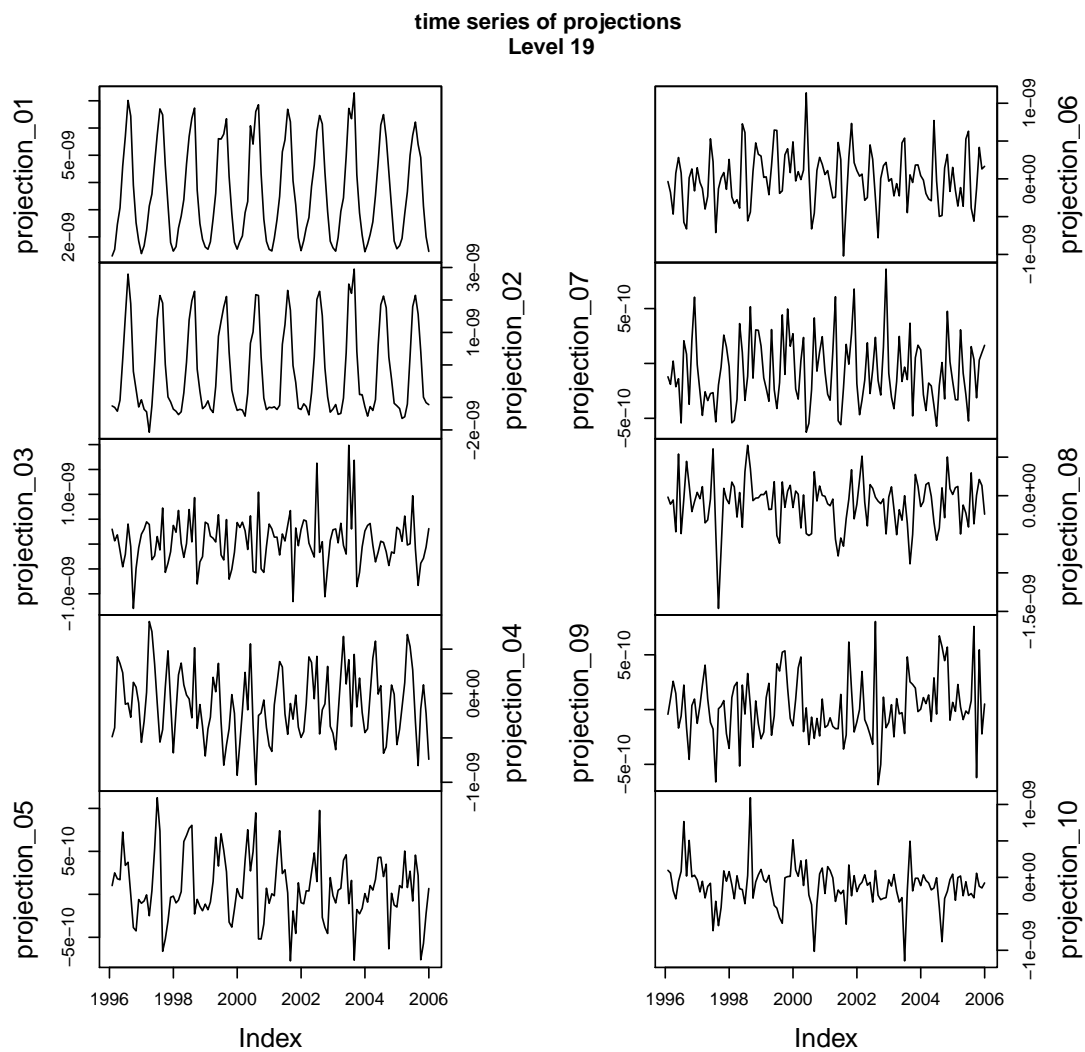


Abb. A.14.: Zeitlicher Verlauf der Hauptkomponenten der PCA der Sulfat-Differenz der untersten Höhenschicht

**Residuen  
der Zeitreihen pro Pixel,  
mit Kovariablen,  
Differenz SO<sub>4</sub>, Level 19**

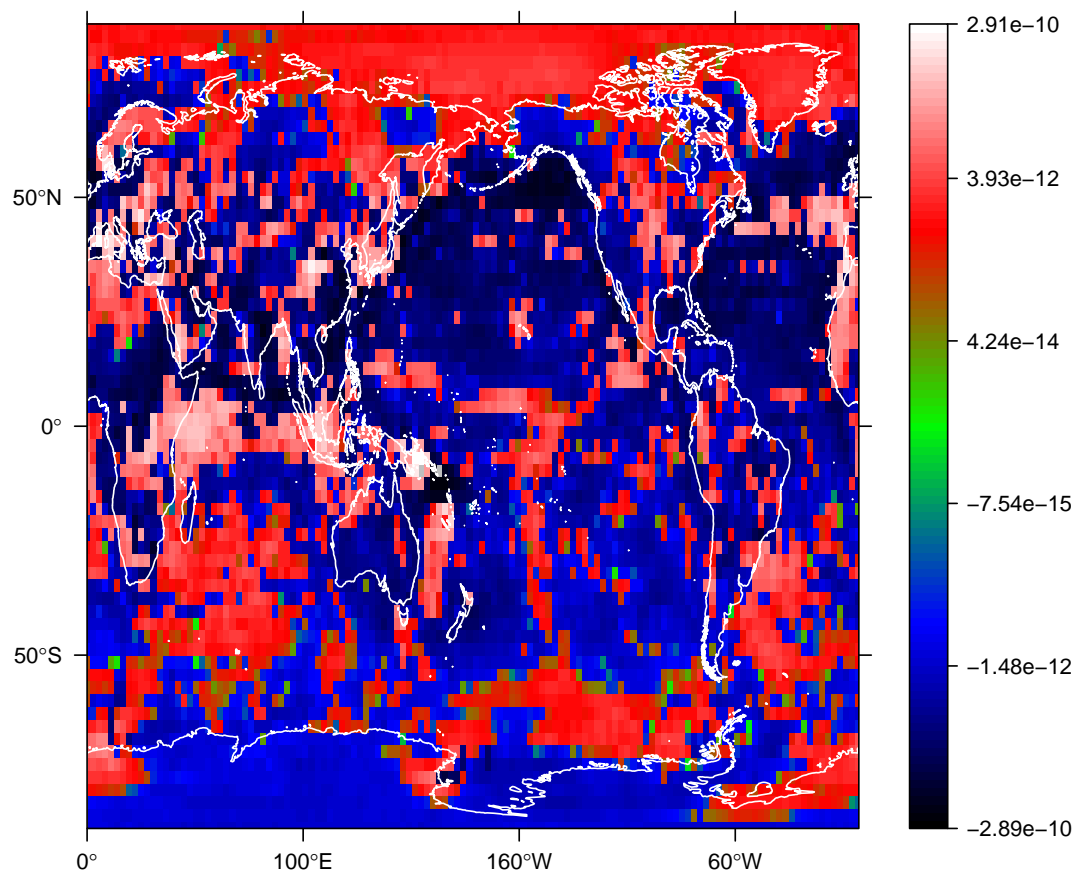


Abb. A.15.: räumliche Verteilung der Residuen des Regressionsmodells pro Pixel aus Kategorie 2 mit Schiffs-, Landemissionen und Niederschlag als Kovariablen für die Sulfat-Konzentration der untersten Höhenschicht

Normal-Q-Q-Plots der  
Zeitreihen mit Kovariablen,  
Differenz SO<sub>4</sub>, Level 19

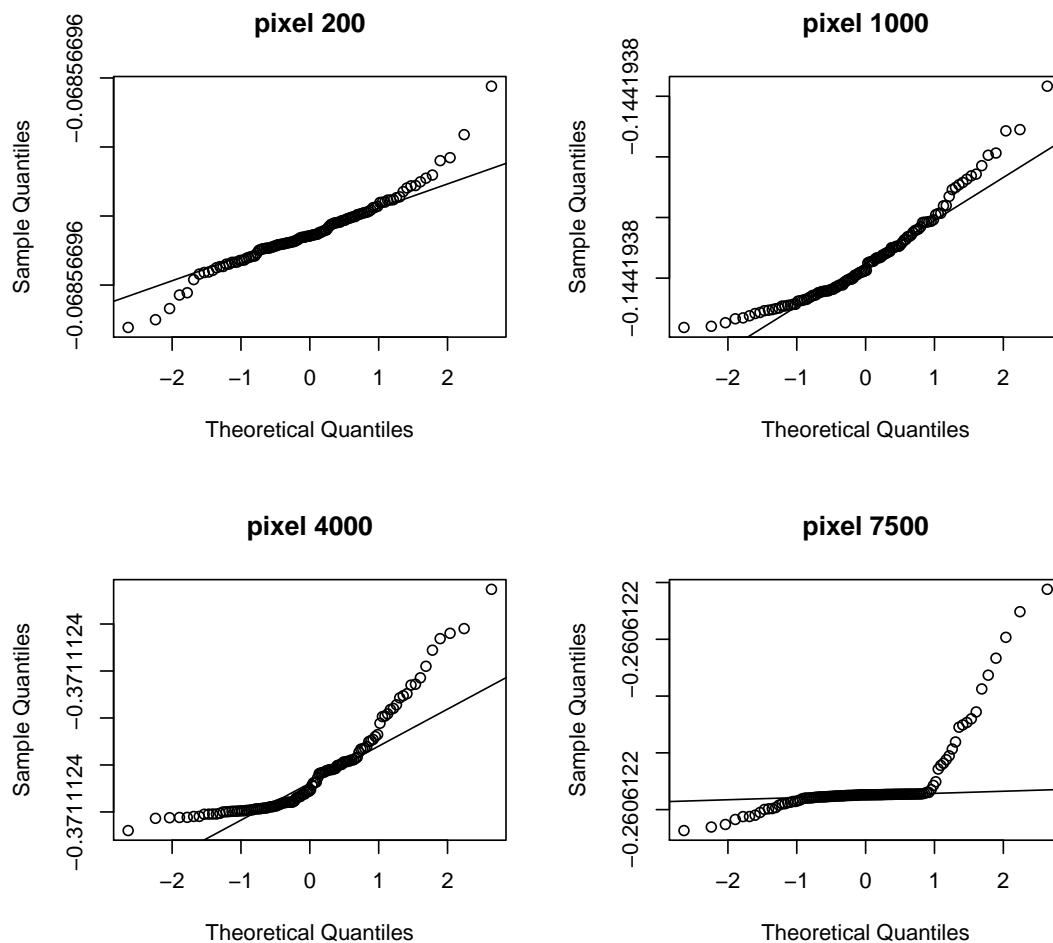


Abb. A.16.: Normal-Q-Q-Plots (zur Untersuchung der Normalverteilungsannahme) der Residuen ausgewählter Pixel des Regressionsmodells pro Pixel aus Kategorie 2 mit Schiffs-, Landemissionen und Niederschlag als Kovariablen für die Sulfat-Konzentration der untersten Höschicht

ACFs der Zeitreihen  
mit Kovariablen,  
Differenz SO4, Level 19

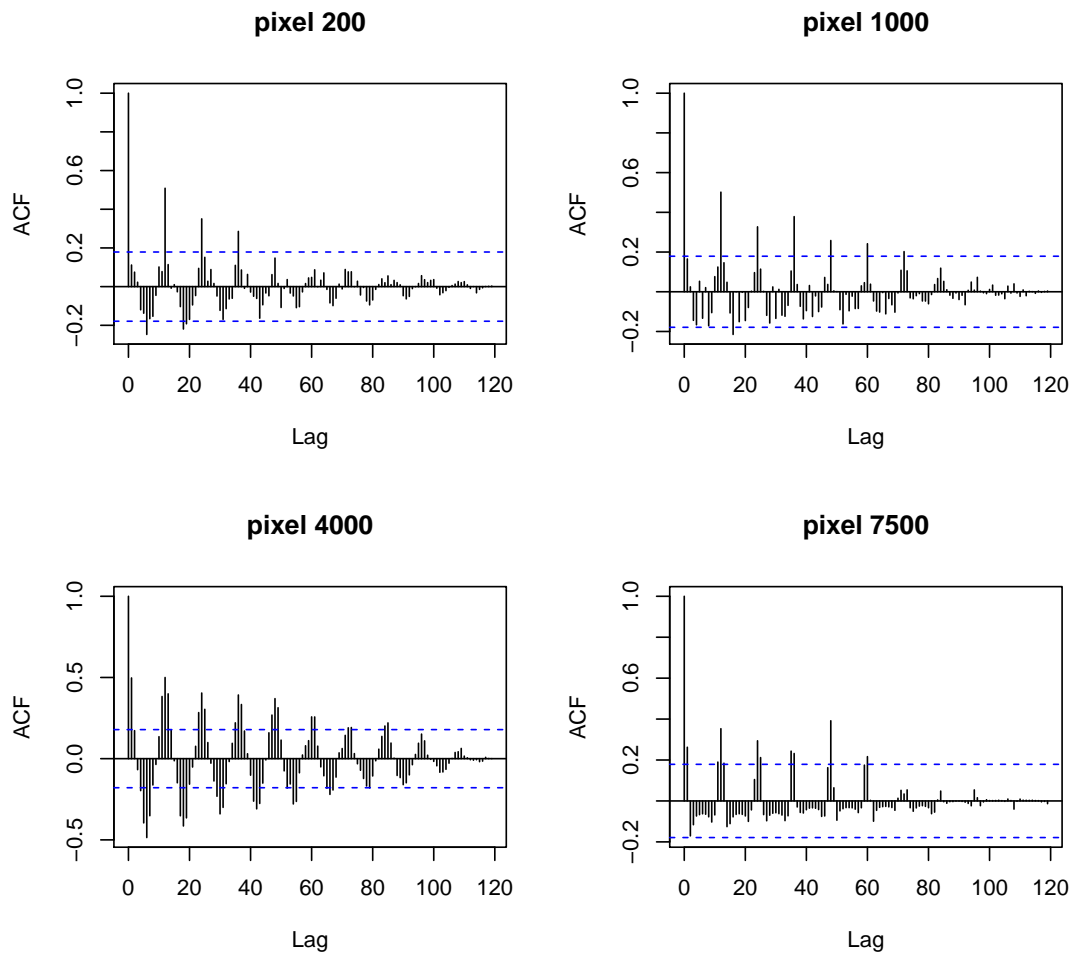


Abb. A.17.: Autokorrelationsfunktion (ACF) der Residuen ausgewählter Pixel des Regressionsmodells pro Pixel aus Kategorie 2 mit Schiffs-, Landemissionen und Niederschlag als Kovariablen für die Sulfat-Konzentration der untersten Höhenschicht

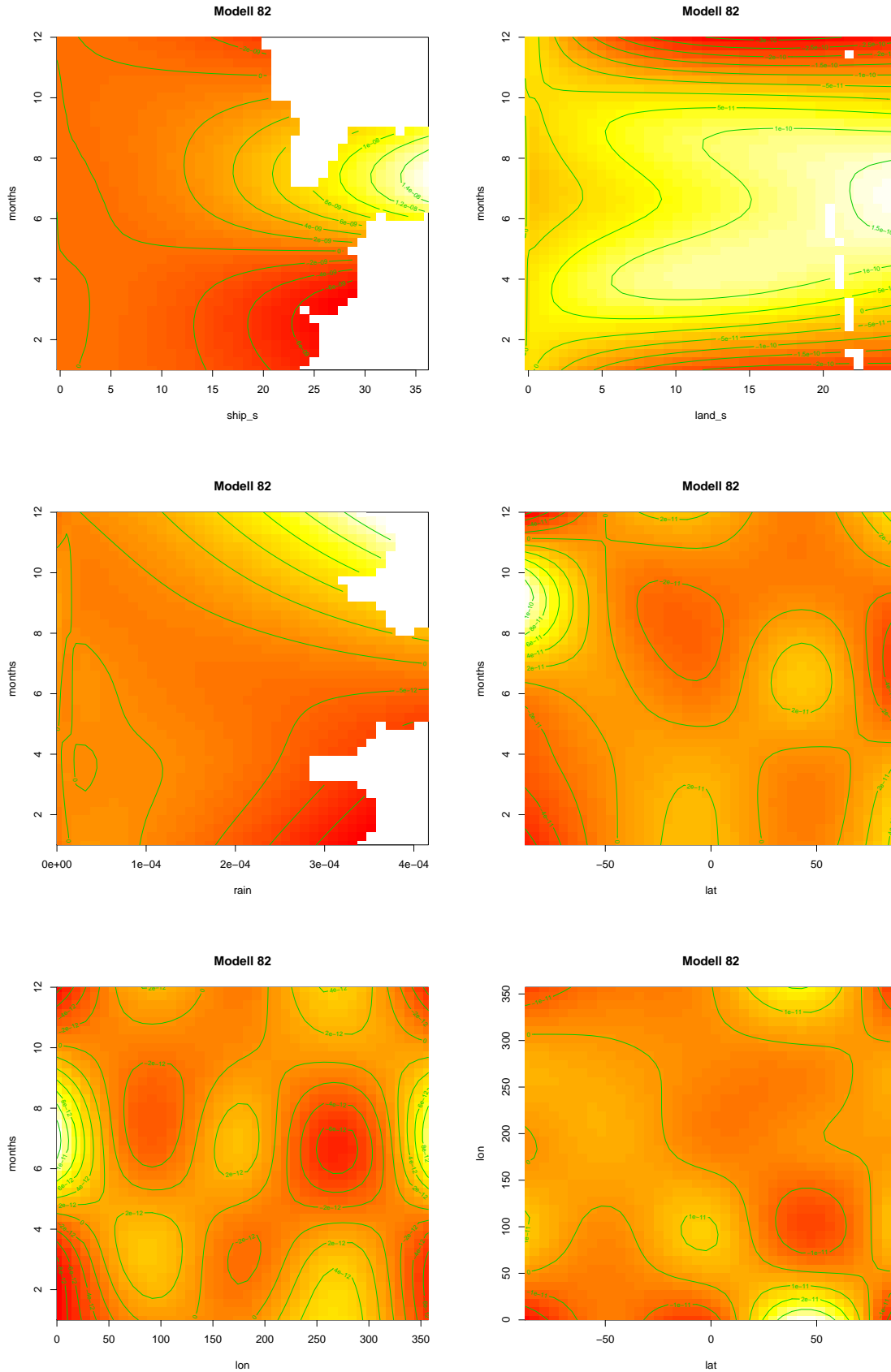


Abb. A.18.: Darstellung einer Auswahl der zweidimensionalen Interaktionen des besten additiven Modells für die Differenz der Sulfat-Konzentration in der untersten Höhenschicht. Dabei ist zu beachten, dass die (durch die Intensität der Farbe) dargestellten Koeffizienten nicht die Koeffizienten an sich, sondern das Produkt aus Wert der Kovariablen und Koeffizient sind. Die (Farb-)Skala der z-Achse entspricht damit der Skala der zu modellierenden Größe (hier den Konzentrationsdifferenzen)

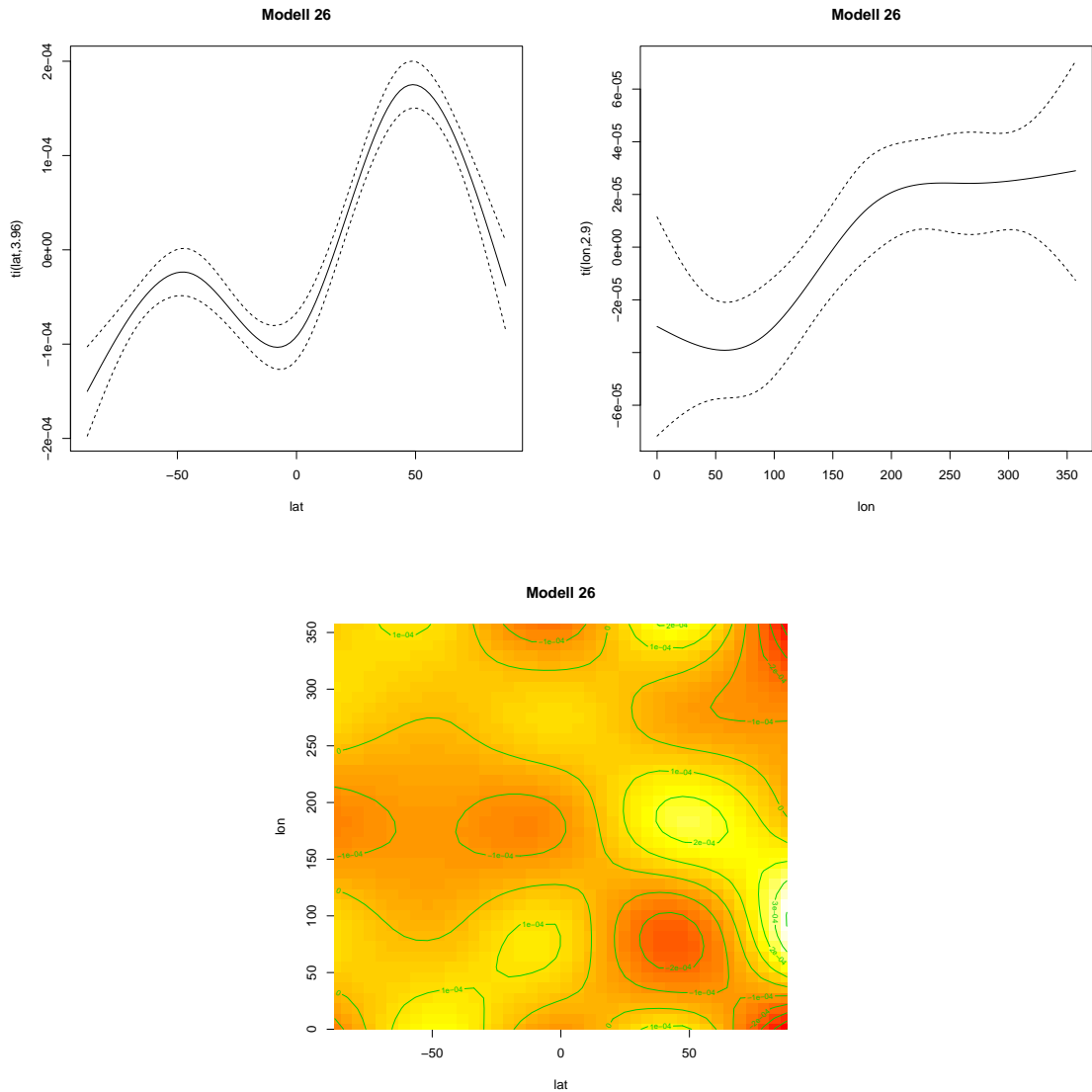


Abb. A.19.: Darstellung der Haupt- und Interaktionseffekte des besten additiven Modells für die Differenz des Wolkenbedeckungsgrades in der untersten Hörschicht. Dabei ist zu beachten, dass die Koeffizienten nicht direkt gegen die Werte der Variablen, sondern das jeweilige Produkt aus Koeffizient und Variablenwert gegen den Wert der Variablen aufgetragen ist. Die Skala der y-Achse beziehungsweise der Farbskala entspricht damit der Skala der zu modellierenden Größe (hier den Differenzen)



# B. Abkürzungsverzeichnis und Liste der verwendeten R-Pakete

## B.1. Abkürzungsverzeichnis

ACF	Autokorrelationsfunktion (AutoCorrelation Function)
AIC	Akaike-Informationskriterium (Akaike Information Criterion)
ARMA	AutoRegressive Moving Average Modell (Zeitreihentheorie)
ARIMA	AutoRegressive Integrated Moving Average Modell (Zeitreihentheorie)
BIC	Bayes-Schwarz-Informationskriterium (Bayesian Information Criterion)
cdo	Climate Data Operators (Algorithmensammlung)
EEOF	Extended EOF Analyse
EOF	Empirical Orthogonal Funktionen (= PCA)
FDR	False-Discovery-Rate (Hypothesentests)
fMRI	functional Magnetic Resonance Imaging
FWER	Family-Wise Error Rate (Hypothesentests)
GAM	generalisierte additive Modelle (Generalized Additive Models)
GAMLSS	generalisierte additive Modelle für Lokations-, Skalen- und Formparameter (Generalized Additive Models for Location, Scale and Shape)
GAMM	gemischte generalisierte additive Modelle (Generalized Additive Mixed Models)
GCM	allgemeine Zirkulationsmodelle (Generalized Circulation Models), Klimamodelle
GLM	generalisierte lineare Modelle (Generalized Linear Models)
GLMM	gemischte generalisiert lineare Modelle (Generalized Linera Mixed Models)
GMRF	Gauß-Markov-Zufallsfelder (Gaussian Markov Random Fields)
ICA	Independend Components Analysis (Alternative zu EOF)
INLA	Integrated Nested Laplace Approximations (Algorithmus zur Berechnung von bayesianischen Modellen)
LASSO	Least Absolute Shrinkage and Selection Operator (Modellwahl)
LMM	gemischte lineare Modelle (Linear Mixed Models)
MARS	Multivariate Adaptive Regression Splines (Alternative zu GAMs)
MM	gemischte Modelle (Mixed Models)
PCA	Hauptkomponentenanalyse (Principal Components Analysis), (= EOF)
QQ-Plot	Quantil-Quantil-Plot zum Vergleichen von Verteilungen
R	Statistiksoftwarepaket

## B.2. Verwendete R-Pakete

automap	[29]
chron	[32]
data.table	[17]
DescTools	[18]
dynlm	[77]
fastICA	[42]
gamlss	[52]
geoR	[34]
gstat	[47]
hydroGOF	[76]
INLA	[39]
inline	[65]
lattice	[58]
maps	[15]
maptools	[10]
MCMCpack	[43]
mgev	[72]
ncdf4	[49]
raster	[30]
sp	[54]
spacetime	[46]
xts	[55]

# C. Elektronischer Anhang

Aufgrund der komplexen Struktur des Datensatzes und der Vielzahl an untersuchten Methoden sind im Rahmen dieser Arbeit viele Grafiken entstanden, die für die allgemeine Auswertung des Datensatzes benötigt wurden, jedoch in Hinblick auf den Erkenntnisgewinn bezüglich der Funktionsweise der Methoden nicht mehr relevant sind. Im Text wurde vor allem die unterste Höhenschicht exemplarisch untersucht. Die gleichen Verfahren wurden auch auf die anderen Höhenschichten angewendet. Diese Auswertungen und die großen Tabellen zur Modellwahl der Regressionsmodelle finden sich im elektronischen Anhang auf der beiliegenden CD. Die dortige Ordnerstruktur folgt in ihrem Aufbau der Struktur der Kapitel des Textes. Zur besseren Orientierung ist im Folgenden eine Übersicht gegeben.

## C.1. Klimamodelle und deskriptive Beschreibung des Datensatzes

### C.1.1. räumliche Verteilung der Emissionen und des Niederschlags

- Anhang/deskriptiv/Kovariablenmuster.pdf: Jahresgang der standardisierten Schiffs- und Landemissionen und des Niederschlags. Zusätzlich Vergleich der verschiedenen geglätteten Muster der Schiffsemissionen im Juli.

### C.1.2. Verteilung der Variablen: Histogramme

- Anhang/deskriptiv/Histogramme.pdf: Histogramme der kompletten (Referenz-Simulation + NoShip-Simulation) Werte, der logarithmierten Werte und der Differenzen

### C.1.3. Veränderung der Variablen mit der Höhenschicht

- Anhang/deskriptiv/Summary\_per\_level.pdf: 5-Punkte-Zusammenfassung (Minimum, erstes Quartil, Median, Mittelwert, drittes Quartil, Maximum) globalen und zeitlichen Mittelwerte pro Höhenschicht der Variablen an sich, ihrer Differenzen und ihrer relativen Differenzen aufgetragen gegen die Höhenschicht (kleine Werte = obere Modellatmosphäre, große Werte = untere Modellatmosphäre)
- Anhang/deskriptiv/deskriptiv\_sumtab.pdf: Tabelle mit 5-Punkte-Zusammenfassung pro Simulation und der Differenz für alle vier Variablen

### C.1.4. Jahresgang und Jahresmittel der Variablen

- Anhang/deskriptiv/Jahresgang\_\*variablename\*.pdf: räumliche Verteilung der Mittelwerte der jeweiligen Monate pro Höhenschicht

- Anhang/deskriptiv/Jahresmittel\_\*variablename\*.pdf: räumliche Verteilung der Jahresmittelwerte pro Höhengschicht

### C.1.5. zeitliche Autokorrelation der Variablen

- Anhang/deskriptiv/ACFglobal\_\*variablename\*.pdf: empirische Autokorrelationsfunktion der globalen Mittelwerte der Variablen und der Differenzen pro Höhengschicht

### C.1.6. Ähnlichkeit der räumlichen Muster zwischen den Höhengschichten

- Anhang/deskriptiv/deskriptiv\_simi.pdf: Pearson-Korrelationskoeffizient und Modified Index of Agreement zwischen den Höhengschichten für das zeitliche Mittel der Differenzen pro Höhengschicht als Maße für die Ähnlichkeit der räumlichen Muster

### C.1.7. Räumliche Korrelation

- Anhang/deskriptiv/lisa\_\*variablename\*.pdf: räumliche Verteilung des lokalen Moran-I's und Geary-C's für das zeitliche Mittel der Variablen und der Differenzen pro Höhengschicht mit unterschiedlicher Größe des Berechnungsradius (3 Pixel, 5 Pixel und 7 Pixel). Je höher der Index, desto größer ist der Einfluss der jeweiligen Region auf die Gesamtstatistik.
- Anhang/deskriptiv/Variogram\_spatial\_\*variablename\*.pdf: Darstellung der Varianz zwischen Punktepaaen mit unterschiedlichem räumlichen Abstand für verschiedene Richtungen ( $0^\circ$  = Nord-Süd,  $90^\circ$  = Ost-West,  $45^\circ$  = NordOst-SüdWest,  $135^\circ$  = NordWest-SüdOst) und omnidirektional sortiert nach absteigenden Höhengschichten (1. Bild = oberste Modellschicht, 19. Bild = unterste Modellschicht). Die Distanz ist in Grad gemessen.

### C.1.8. Raum-zeitliche Korrelation

- Anhang/deskriptiv/Variogram\_spatiotemporal\_\*variablename\*.pdf: Darstellung der Varianz zwischen Punktepaaen mit unterschiedlichem raum-zeitlichem Abstand in zwei verschiedenen Darstellungen für die Variablen und die Differenzen pro Höhengschicht.

## C.2. Hypothesentests

In sämtlichen Grafiken ist der jeweilige Test zum Signifikanzniveau  $\alpha = 0.05$  durchgeführt. Türkis sind p-Werte kleiner als  $\alpha$ , welche zu einer Ablehnung der Nullhypothese führen. Rot sind die p-Werte markiert, welche die Nullhypothese nicht ablehnen. Die jeweiligen Einzelbilder stellen die Ergebnisse unter Verwendung von multiplen Testkorrekturen dar: none = keine Korrektur, BH = Benjamini-Hochberg, BY = Benjamini-Yekutieli, hommel = Hommel, bonferroni = einfache

Bonferroni-Korrektur, Holm = Holm's Stepwise Prozedur, Hochberg = Hochberg's Stepwise Prozedur.

### C.2.1. Shapiro-Wilk-Tests

- Anhang/hypothesen/Tests\_\*variablename\*\_swtest\_19: p-Werte des Shapiro-Wilk-Tests für die Jahresmittel und jeweiligen Monate der Differenzen für die unterste Höenschicht
- Anhang/hypothesen/Tests\_\*variablename\*\_swtest\_abs\_19: p-Werte des Shapiro-Wilk-Tests für die Jahresmittel und jeweiligen Monate des Absolutbetrages der Differenz für die unterste Höenschicht

### C.2.2. Welch-t-Tests

- Anhang/hypothesen/Tests\_\*variablename\*\_ttest\_twoside\_19: p-Werte des zweiseitigen Welch-t-Tests für die Jahresmittel und jeweiligen Monate der Differenzen für die unterste Höenschicht
- Anhang/hypothesen/Tests\_\*variablename\*\_ttest\_oneside\_19: p-Werte des einseitigen Welch-t-Tests für die Jahresmittel und jeweiligen Monate des Absolutbetrags der Differenzen für die unterste Höenschicht mit der Schranke 0
- Anhang/hypothesen/Tests\_\*variablename\*\_ttest\_oneside\_decimal\_19: p-Werte des Welch-t-Tests für die Jahresmittel und jeweiligen Monate des Absolutbetrags der Differenzen für die unterste Höenschicht mit der zweiten Nachkommastelle des globalen Mittels als Schranke
- Anhang/hypothesen/Tests\_\*variablename\*\_ttest\_oneside\_quantile\_19: p-Werte des Welch-t-Tests für die Jahresmittel und jeweiligen Monate des Absolutbetrags der Differenzen für die unterste Höenschicht mit dem 25%-Quantil als Schranke

### C.2.3. Wilcox-Vorzeichen-Rang-Tests

- Anhang/hypothesen/Tests\_\*variablename\*\_ttest\_twoside\_19: p-Werte des zweiseitigen Wilcox-Tests für die Jahresmittel und jeweiligen Monate der Differenzen für die unterste Höenschicht
- Anhang/hypothesen/Tests\_\*variablename\*\_ttest\_oneside\_19: p-Werte des einseitigen Wilcox-Tests für die Jahresmittel und jeweiligen Monate des Absolutbetrags der Differenzen für die unterste Höenschicht mit der Schranke 0
- Anhang/hypothesen/Tests\_\*variablename\*\_ttest\_oneside\_decimal\_19: p-Werte des Wilcox-Tests für die Jahresmittel und jeweiligen Monate des Absolutbetrags der Differenzen für die unterste Höenschicht mit der zweiten Nachkommastelle des globalen Mittels als Schranke

- Anhang/hypothesen/Tests\_ \*variablename\* \_ttest\_oneside\_quantile\_19: p-Werte des Wilcoxon-Tests für die Jahresmittel und jeweiligen Monate des Absolutbetrags der Differenzen für die unterste Höhenschicht mit dem 25%-Quantil als Schranke

#### C.2.4. Tabellen mit der Anzahl der abgelehnten Nullhypothesen

- Anhang/hypothesen/Anzahl\_abgelehnte\_H0\_alpha5perc.pdf: Für alle oben grafisch dargestellten Tests und ihren multiplen Testkorrekturen die Anzahl der abgelehnten Nullhypothesen in Prozent pro Höhenschicht und Variable

### C.3. PCA/EOF-Analyse und ICA

#### C.3.1. Vergleich der Algorithmen cdo und fastICA in R

- Anhang/pcaeof/Vergleich\_R\_cdo\_ \*variablename\*.pdf: Vergleich der Eigenvektoren für die PCA der unveränderten, zentrierten und standardisierten Jahresmittelwerte der Differenz

#### C.3.2. Eigenwerte, Eigenvektoren und Hauptkomponenten der PCA

- Anhang/pcaeof/EOF\_ \*zentriert\* \_ \*variablename\* \_ \*zeitlich\*.pdf: Eigenwerte, Eigenvektoren und Hauptkomponenten der PCA pro Höhenschicht. Dabei bedeutet der Wert von \*zentriert\*: nochange = unveränderte Werte, centered = zentrierte Werte, scaled = standardisierte Werte und \*zeitlich\* heißt: allyears = kein zeitliche Mittel, yearmean = Jahresmittelwerte, mon1-12: nur die Werte der jeweiligen Monate (alle Januare, alle Februar, ...)
- Anhang/pcaeof/pcaeof\_hotelling\_ \*multtest\* \_ \*variablename\*: p-Werte eines Hotelling- $T^2$ -Tests mit variabel vielen Hauptkomponenten. \*multtest\* = benjyekut gibt an, dass die p-Werte mit der multiplen-Test-Korrektur von Benjamini-Yekutieli nachbearbeitet wurden.

#### C.3.3. Eigenvektoren und Hauptkomponenten der ICA

- Anhang/pcaeof/ICA\_EigVec\_ \*variablename\* \_ \*zeitlich\* \_ \*level\*.pdf: Eigenvektoren der ICA pro Höhenschicht \*level\* mit den unveränderten, zentrierten und skalierten Variablen. Für \*zeitlich\* = yearmean wurden die Jahresmittelwerte verwendet, ansonsten keine zeitliche Mittelung durchgeführt.
- Anhang/pcaeof/ICA\_Proj\_ \*variablename\* \_ \*zeitlich\* \_ \*level\*.pdf: Hauptkomponenten der ICA pro Höhenschicht \*level\* mit den unveränderten, zentrierten und skalierten Variablen. Für \*zeitlich\* = yearmean wurden die Jahresmittelwerte verwendet, ansonsten keine zeitliche Mittelung durchgeführt.

## C.4. Regressionsmodelle

### C.4.1. Modellgüte

- Anhang/regression/Modellguete\_*\*variablename\**.pdf: Liste der verschiedenen Modelle und ihre jeweiligen zugehörigen Modellgütekriterien AIC, BIC und  $R^2$  (allgemein als Anteil der erklärten Devianz bezeichnet). AIC und BIC sind als Differenz zum minimalen Wert (bestes Modell) aller Modelle (linear und additiv) angegeben. Die Spalte Interact gibt an, ob Interaktionen zwischen den im Modell enthaltenen Kovariablen ebenfalls enthalten waren oder nicht. Die darauffolgenden Spalten geben jeweils an, ob die Kovariable im Modell enthalten war: lat = Breitengrad, lon = Längengrad, months = Monate (1 bis 12), years = Jahre (1996 bis 2005), time = fortlaufender Zeitindex (1 bis 120), ship/land/rain = Schiffs- bzw. Landemissionen und Niederschlag (eventuelle Glättung ist durch den Namen des verwendeten Glättungsfilters gekennzeichnet, bin = gerundete Version der standardisierten Emissionen), special = Spezielle Eigenschaften des Modells (bestoff = Kombination der am besten passenden Einzelvariablen der linearen Modelle, factor = Breiten- und Längengrad als diskrete Variable, HigherResolution = höhere Auflösung der Spline-Basis-Knoten, CyclicalBasis = zyklische Basisfunktionen für Längengrade oder Monate, TransformedLatLon = Modelle mit  $lat^3$  um die Pole schwächer zu gewichten, KrigingBasis = Basisfunktion für die räumlichen Variablen, welche dem Kriging entsprechen).

### C.4.2. Lineare Modelle

- Anhang/regression/Linreg\_all\_bestof\_coefs\_*\*variablename\**.pdf: Tabelle mit den Koeffizienten und den jeweiligen dazugehörigen t-Test-Ergebnissen des besten linearen Modells
- Anhang/regression/Linreg\_all\_bestof\_diagnosis\_*\*variablename\**.png: Hier sind vier Diagnoseplots des besten Modells zu sehen. Der erste Plot oben links ist ein Normal-Q-Q-Plot worin die Verteilung der Residuen gegen eine Normalverteilung aufgetragen ist. Im zweiten Plot oben links sind die Residuen gegen die vom Modell berechneten Werte aufgetragen. Unten links ist ein Histogramm der Residuen zu sehen und unten rechts sind die wahren Werte gegen die berechneten aufgetragen.

### C.4.3. GLMs und GAMLSS

- Anhang/regression/glms\_cov\_diff.pdf und Anhang/regression/glms\_wocov\_diff.pdf: Diagnoseplots und Koeffizientenplots zum Vergleich der jeweils durch eine Backward-Selection als am besten ausgewählten Modelle für zwei verschiedene GLMs (gaussian\_id: Normalverteilung + keine Linkfunktion, gaussian\_inv: Normalverteilung + inverse Linkfunktion) für die Differenz der Sulfat-Konzentration. Diese beiden waren aus einer Kombination von 3 Verteilungen (Normal, Gamma, invers-Normal) und 3 Linkfunktionen (Identität, Logarithmus und Invers) die einzigen Kombinationen welche nicht zu einem Abbruch des Algorithmus geführt haben. In

der *\*cov\**-Variante wurden dabei die Emissions- und Niederschlagsvariable als Kovariablen verwendet, in der *\*wocov\**-Variante die Koordinaten.

- Anhang/regression/Alternativen\_gamlss\_jsu.pdf und Anhang/regression/Alternativen\_gamlss\_shasho.pdf: Diagnoseplots für lineare Modelle mit zwei verschiedenen Verteilungsannahmen (jsu = Johnson's SU, shasho = sinh-arcsinh) für die Sulfat-Differenz in der untersten Höhengschicht. Es konnten nur ein Monat modelliert werden, da die Rechenzeit für alle Daten zu lange war. Als Kovariablen wurden die durch Rundung diskretisierten Emissions- und Niederschlags-Variablen verwendet.

#### C.4.4. Zeitliche lineare Modelle pro Pixel

- Anhang/regression/Linreg\_proPixel.pdf: Koeffizienten- und Diagnoseplots für jeweils ein Beispielmmodell beiden Kategorien (Erklärung sh. Abschnitt ??) für zeitliche Modelle der Sulfat-Differenz der untersten Höhengschicht pro Pixel
- Diagnoseplots weiterer zeitlicher Regressionsmodelle pro Pixel für die Sulfat-Differenz der untersten Höhengschicht:
  - Anhang/regression/pixel\_diff\_res\*.pdf: Räumliche Verteilung der Residuen
  - Anhang/regression/pixel\_diff\_res\_acf\*.pdf: Autokorrelationsfunktion (ACF) und Normal-Q-Q-Plot (zur Überprüfung der Normalverteilungsannahme) der Residuen für vier beispielhaft ausgewählte Pixel
  - Anhang/regression/pixel\_diff\_rsqr\*.pdf: räumliche Verteilung des Modellgütemaßes  $R^2$

#### C.4.5. Lineare Modelle pro Zeitpunkt

- Anhang/regression/Linreg\_proZeit\_\*.variablename\*.pdf: zeitlicher Verlauf der Modellkoeffizienten für Modell 1,2 und 3 (Beschreibung in Abschnitt 5.2.1.4) für die unterste Höhengschicht

#### C.4.6. Kriging-Modelle pro Zeitpunkt

- Anhang/regression/Geostat\_pred\_\*.variablename\*.pdf: Koeffizienten- und Diagnoseplots für Kriging-Modelle mit zwei verschiedenen Trend-Modellen für die Differenzen der Konzentrationen und des Wolkenbedeckungsgrades vom Juli 2001. Für das erste Modell wurde ein konstanter Trend angenommen, das zweite Modell enthält die Emissions- und Niederschlagsvariablen als Kovariablen. Die jeweils erste Grafik vergleicht die tatsächlichen Werte mit denen vom Modell vorhergesagten. Der zweite Plot zeigt die räumliche Verteilung der Residuen und der dritte Plot vergleicht die nur vom geschätzten Trend vorhergesagten Modelle mit den von einem linearen Modell vorhergesagten Werten.



### C.4.7. Additive Modelle

- Anhang/regression/Addreg\_coefs\_\*variablename\*.pdf: Darstellung der glatten Funktion pro Kovariable und Interaktion für alle Modelle der Differenzen des untersten Höhenlevels. Dabei ist zu beachten, dass die Koeffizienten nicht direkt gegen die Werte der Variablen, sondern das jeweilige Produkt aus Koeffizient und Variablenwert gegen den Wert der Variablen aufgetragen ist. Die Skala der y-Achse entspricht damit der Skala der zu modellierenden Größe (hier den Differenzen).
- Anhang/regression/Addreg\_resids\_\*variablename\*.pdf: Vier Diagnoseplots pro Modell: räumliche Verteilung der Residuen für den Juli 1996 zur Überprüfung auf verbleibende räumliche Korrelation, Normal-Q-Q-Plots der Residuen ausgewählter Pixel und Monate zur Überprüfung der Normalverteilungsannahme, Autokorrelationsfunktion der Residuen ausgewählter Pixel zur Überprüfung auf verbleibende zeitliche Korrelation, Vergleich der geschätzten Werte mit den tatsächlichen Werten im Juli 1996
- Anhang/regression/Addreg\_pval\_\*variablename\*.pdf: Tabellen mit dem geschätzten Interzept und seinem p-Wert und den p-Werten und geschätzten Freiheitsgraden (kann auch als Maß für die Glattheit verwendet werden) der additiv modellierten Kovariablen und Interaktionen der additiven Modelle für die Differenzen in der untersten Höhengschicht.

### C.4.8. Bayesianische Regressionsmodelle

- Anhang/regression/Alternativen\_mcmc\_linear.pdf: Verteilungsdichte der geschätzten Koeffizienten für ein bayesianisches lineares Regressionsmodell der Differenz der Sulfat-Konzentration in der untersten Höhengschicht. Es wurden die durch Rundung diskretisierten Varianten der Emissions- und Niederschlagsmuster als Kovariablen verwendet. Es sind jeweils die Posteriori-Verteilungsdichten der geschätzten Koeffizienten gezeigt. Die x-Achse ist somit der Wert des Koeffizienten und die y-Achse den Wert der Dichte (Integration über den gesamten Wertebereich ergibt 0). Die Interpretation des Mittelwerts (manchmal auch des Medians) der Verteilung entspricht damit der des Punktschätzer, also dem Wert des Koeffizienten, im linearen Modell. Die Wahrscheinlichkeit, dass der Wert des Koeffizienten in einem bestimmten Bereich liegt, kann somit einfach über Integration der jeweiligen Verteilungsdichte berechnet werden.
- Anhang/regression/Alternativen\_mcmc\_inla.pdf: Koeffizientenplots und Diagnoseplots für zwei verschiedene semi-bayesianische Regressionsmodelle für die standardisierte Differenz der Sulfat-Konzentration in der untersten Höhengschicht über dem Nordpazifik für das Jahr 1996. Das erste Modell verwendet einen zweidimensionalen Random-Walk zur Modellierung der räumlichen Abhängigkeiten, das zweite Modell einen Gaus'schen Prozess mit Matern-Kovarianzfunktion. Die jeweils erste Abbildung zeigt den Jahresgang des Mittelwerts der Posteriori-Verteilung. Die darauffolgenden Abbildungen vergleichen die wahren Werte mit den Posteriori-Mittelwert und zeigen die räumliche Verteilung der Residuen für Februar und Juli.

- Anhang/regression/Alternativen\_mcmc\_inla\_\*.pdf: Verteilungsdichten der geschätzten Koeffizienten für die beiden semi-bayesianischen Regressionsmodelle. \*matern\* steht dabei für das Modell mit der Matern-Kovarianzfunktion, \*rw\* für das Modell mit dem Random-Walk.

## D. Literaturverzeichnis

- [1] *Handbook of Linear Algebra*. Taylor & Francis Inc, 2013.
- [2] Fiaz Ahmad and Rehan Ahmad Khan. A power comparison of various normality tests. *Pak.j.stat.oper.res.*, 11(3):331, sep 2015.
- [3] Abdullah Almutairi and Timothy A. Warner. Change detection accuracy and image properties: A study using simulated data. *Remote Sensing*, 2(6):1508–1529, Jun 2010.
- [4] Luc Anselin. Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2):93–115, sep 2010.
- [5] John Ashburner and Karl J. Friston. Voxel-based morphometry—the methods. *NeuroImage*, 11(6):805–821, June 2000.
- [6] Klaus Backhaus, Bernd Erichson, Wulff Plinke, and Rolf Weiber. *Multivariate Analysemethoden*. Springer Berlin Heidelberg, 2000.
- [7] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.
- [8] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188, 08 2001.
- [9] L Mark Berliner, Richard A Levine, and Dennis J Shea. Bayesian climate change assessment. *Journal of Climate*, 13(21):3805–3820, 2000.
- [10] Roger Bivand and Nicholas Lewin-Koh. *maptools: Tools for reading and handling spatial objects*, 2014. R package version 0.8-30.
- [11] F. DuBois Bowman, Brian Caffo, Susan Spear Bassett, and Clinton Kilts. A Bayesian hierarchical framework for spatial modeling of fmri data. *NeuroImage*, 39(1):146–156, Jan 2008.
- [12] D. M. Bramich. A new algorithm for difference image analysis. February 2008.
- [13] Luca Caucci, Harrison H Barrett, and Jeffrey J Rodriguez. Spatio-temporal hotelling observer for signal detection from image sequences. *Optics express*, 17(13):10946–10958, 2009.
- [14] Cdo 2014: Climate data operators.
- [15] Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka <tpminka@media.mit.edu>. *maps: Draw Geographical Maps*, 2014. R package version 2.3-9.
- [16] Noel Cressie and Christopher K. Wikle. *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Ltd., 2011.

- [17] M Dowle, T Short, S Lianoglou, A Srinivasan with contributions from R Saporita, and E Antonyan. *data.table: Extension of data.frame*, 2014. R package version 1.9.4.
- [18] Andri Signorell et mult. al. *DescTools: Tools for Descriptive Statistics*, 2015. R package version 0.99.15.
- [19] Julian J. Faraway. *Linear Models with R*. Chapman & Hall/CRC, 2005.
- [20] Jerome H. Friedman. Multivariate adaptive regression splines. *Ann. Statist.*, 19(1):1–67, mar 1991.
- [21] Alan E. Gelfand. Hierarchical modeling for spatial data problems. *Spatial Statistics*, 1:30–39, May 2012.
- [22] Eric Gilleland, David Ahijevych, Barbara G. Brown, Barbara Casati, and Elizabeth E. Ebert. Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, 24(5):1416–1430, Oct 2009.
- [23] Sander Greenland and Charles Poole. Living with p values. *Epidemiology*, 24(1):62–68, jan 2013.
- [24] Stefan Güss, Martin Heimann, and Hans von Storch. *Das Klimasystem und seine Modellierung*. Springer, 1999.
- [25] Klaus Hasselmann. Optimal fingerprints for the detection of time-dependent climate change. *Journal of Climate*, 6(10):1957–1971, 1993.
- [26] Klaus Hasselmann. Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics*, 13(9):601–611, 1997.
- [27] G.C. Hegerl, F. W. Zwiers, P. Braconnot, N.P. Gillett, Y. Luo, J.A. Marenco Orsini, N. Nicholls, J.E. Penner, and P.A. Stott. Understanding and attributing climate change. 2007.
- [28] J. Hendricks, B. Kärcher, and U. Lohmann. Effects of ice nuclei on cirrus clouds in a global climate model. *J. Geophys. Res.*, 116(D18), sep 2011.
- [29] P.H. Hiemstra, E.J. Pebesma, C.J.W. Twenhöfel, and G.B.M. Heuvelink. Real-time automatic interpolation of ambient gamma dose rates from the dutch radioactivity monitoring network. *Computers & Geosciences*, 2008. DOI: <http://dx.doi.org/10.1016/j.cageo.2008.10.011>.
- [30] Robert J. Hijmans. *raster: Geographic Data Analysis and Modeling*, 2015. R package version 2.5-2.
- [31] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4-5):411–430, jun 2000.
- [32] David James and Kurt Hornik. *chron: Chronological Objects which Can Handle Dates and Times*, 2014. R package version 2.3-45. S original by David James, R port by Kurt Hornik.

- [33] Devin S Johnson and Jennifer A Hoeting. Bayesian multimodel inference for geostatistical regression models. *PloS one*, 6(11):e25677, 2011.
- [34] Paulo J. Ribeiro Jr and Peter J. Diggle. *geoR: Analysis of Geostatistical Data*, 2015. R package version 1.7-5.1.
- [35] Sadanori Konishi. *Introduction to Multivariate Analysis*. Taylor & Francis Inc, 2014.
- [36] Yufeng Kou. *Abnormal Pattern Recognition in Spatial Data*. PhD thesis, Virginia Polytechnic Institute and State University, November 2006.
- [37] Christine Land, Michael Ponater, Robert Sausen, and Erich Roeckner. The echam4.l39(dlr) atmosphere gcm - technical description and model climatology. Technical report, DLR-Institut für Physik der Atmosphäre and Max-Planck-Institut für Meteorologie, 1999.
- [38] Stephen S Leroy. Detecting climate signals: Some bayesian aspects. *Journal of Climate*, 11(4):640–651, 1998.
- [39] Finn Lindgren and Håvard Rue. Bayesian spatial modelling with r - INLA. *Journal of Statistical Software*, 63(19), 2015.
- [40] Martin A Lindquist and Andrew Gelman. Correlations and multiple comparisons in functional imaging: A statistical perspective (commentary on vul et al., 2009). *Perspectives on Psychological Science*, 4(3):310–313, 2009.
- [41] Murat İlsever and Cem Ünsalan. *Pixel-Based Change Detection Methods*. Springer Science + Business Media, 2012.
- [42] J L Marchini, C Heaton, and B D Ripley. *fastICA: FastICA Algorithms to perform ICA and Projection Pursuit*, 2013. R package version 1.2-0.
- [43] Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park. MCMCpack: Markov chain monte carlo in R. *Journal of Statistical Software*, 42(9):22, 2011.
- [44] Seung-Ki Min, A Hense, H Paeth, WT Kwon, and JH Oh. 1.4 a general bayesian approach to climate change detection and attribution using echam3-lsg scenario runs and ncep/ncar reanalysis observations. 2000.
- [45] Thomas Nichols and Satoru Hayasaka. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, 12(5):419–446, Oct 2003.
- [46] Edzer Pebesma. spacetime: Spatio-temporal data in r. *J. of Stat. Software*, 51(7), 2012.
- [47] Edzer J. Pebesma. Multivariable geostatistics in s: the gstat package. *Computers & Geosciences*, 30:683–691, 2004.
- [48] Francisco Pereira, Tom Mitchell, and Matthew Botvinick. Machine learning classifiers and fmri: A tutorial overview. *NeuroImage*, 45(1):S199–S209, Mar 2009.

- [49] David Pierce. *ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files*, 2015. R package version 1.15.
- [50] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [51] Richard J Radke, Srinivas Andra, Omar Al-Kofahi, and Badrinath Roysam. Image change detection algorithms: a systematic survey. *Image Processing, IEEE Transactions on*, 14(3):294–307, 2005.
- [52] R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, 54:507–554, 2005.
- [53] M. Righi, J. Hendricks, and R. Sausen. The global impact of the transport sectors on atmospheric aerosol: simulations for year 2000 emissions. *Atmos. Chem. Phys.*, 13(19):9939–9970, 2013.
- [54] Virgilio Gomez-Rubio Roger S. Bivand, Edzer Pebesma. *Applied Spatial Data Analysis with R*. Springer-Verlag GmbH, 2013.
- [55] Jeffrey A. Ryan and Joshua M. Ulrich. *xts: eXtensible Time Series*, 2014. R package version 0.9-7.
- [56] Sujit Kumar Sahu and Khandoker Shuvo Bakar. Hierarchical Bayesian autoregressive models for large space-time data with applications to ozone concentration modelling. *Applied Stochastic Models in Business and Industry*, 28(5):395–415, Sep 2012.
- [57] Stephan R. Sain, Reinhard Furrer, and Noel Cressie. A spatial analysis of multivariate output from regional climate models. *The Annals of Applied Statistics*, 5(1):150–175, Mar 2011.
- [58] Deepayan Sarkar. *Lattice: Multivariate Data Visualization with R*. Springer, New York, 2008. ISBN 978-0-387-75968-5.
- [59] W. Sauerbrei, P. Royston, and H. Binder. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Statistics in Medicine*, 26(30):5512–5528, 2007.
- [60] David A. Savitz and Andrew F. Olshan. Multiple comparisons and related issues in the interpretation of epidemiologic data. *American Journal of Epidemiology*, 142(9):904–908, 1995.
- [61] R. Schnur and Kl. Hasselmann. Optimal filtering for Bayesian detection and attribution of climate change. *Climate Dynamics*, 24(1):45–55, Dec 2004.
- [62] J P Shaffer. Multiple hypothesis testing. *Annual Review of Psychology*, 46(1):561–584, 1995.
- [63] Xiaotong Shen, Hsin-Cheng Huang, and Noel Cressie. Nonparametric hypothesis testing for a spatial signal. *Journal of the American Statistical Association*, 97(460):1122–1140, Dec 2002.

- [64] Matt Silver, Giovanni Montana, and Thomas E. Nichols. False positives in neuroimaging genetics using voxel-based morphometry data. *NeuroImage*, 54(2):992–1000, Jan 2011.
- [65] Oleg Sklyar, Duncan Murdoch, Mike Smith, Dirk Eddelbuettel, Romain Francois, and Karline Soetaert. *inline: Functions to Inline C, C++, Fortran Function Calls from R*, 2015. R package version 0.3.14.
- [66] John R. Thompson. Invited commentary: Re: Multiple comparisons and related issues in the interpretation of epidemiologic data. *American Journal of Epidemiology*, 147(9):801–806, 1998.
- [67] Helge Toutenburg, Christian Heumann, Malte Wißmann, and Michael Schomaker. *Induktive Statistik*. Springer-Verlag GmbH, 2008.
- [68] H. von Storch and A. Navarra, editors. *Analysis of Climate Variability*. Springer, 1999.
- [69] Hans von Storch. *Statistical Analysis in Climate Research*. CAMBRIDGE UNIVERSITY PRESS, 2002.
- [70] Cort J. Willmott, Steven G. Ackleson, Robert E. Davis, Johannes J. Feddema, Katherine M. Klink, David R. Legates, James O’Donnell, and Clinton M. Rowe. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, 90(C5):8995–9005, 1985.
- [71] Simon N. Wood. *Generalized Additive Models, An Introduction with R*. Chapman & Hall/CRC, 2006.
- [72] Simon N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36, 2011.
- [73] Mark W Woolrich, Mark Jenkinson, J Michael Brady, and Stephen M Smith. Fully bayesian spatio-temporal modeling of fmri data. *Medical Imaging, IEEE Transactions on*, 23(2):213–231, 2004.
- [74] K. J. Worsley, A. C. Evans, S. Marrett, and P. Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918, nov 1992.
- [75] B. W. Yap and C. H. Sim. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155, dec 2011.
- [76] Mauricio Zambrano-Bigiarini. *hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series*, 2014. R package version 0.3-8.
- [77] Achim Zeileis. *dynlm: Dynamic Linear Regression*, 2014. R package version 0.3-3.

## **E. Danksagung und Eigenständigkeitserklärung**

Ich möchte mich an dieser Stelle ganz herzlich bei allen beteiligten Personen bedanken, die es mir ermöglicht haben, diese Arbeit durchzuführen. An vorderster Stelle danke ich meinem Betreuer Herrn Dr. Michael Ponater vom Institut für Physik der Atmosphäre des Deutschen Zentrums für Luft- und Raumfahrt welcher mir mit Rat und Tat beiseite stand und mit hilfreichen Kommentaren Ordnung in meine wirre Logik gebracht hat. Ebenso bedanke ich mich bei Herrn Prof. Dr. Volker Schmid vom Institut für Mathematik, Informatik und Statistik der Ludwig-Maximilians-Universität München und Herrn Prof. Dr. Robert Sausen ebenfalls vom DLR dafür, dass sie die passenden Rahmenbedingungen für die Durchführung der Arbeit geschaffen haben. Außerdem gilt ein großes Dankeschön auch den Entwicklern und Programmierern, welche ihre R-Erweiterungen öffentlich zugänglich gemacht haben, ohne welche die Durchführung dieser Analyse wesentlich länger gedauert hätte. Widmen möchte ich diese Arbeit meinem Mann Michael, als Dank für die Geduld und Unterstützung während dieser anstrengenden Zeit.



Hiermit erkläre ich, dass die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden.

München, den ....., .....